

Bauhaus-Universität Weimar
Fakultät Medien
Studiengang Medieninformatik

Automatische Erkennung von Clickbait auf Twitter

Bachelorarbeit

Sebastian Köpsel
geb. am: 01.01.1990 in Neubrandenburg

Matrikelnummer 100146

1. Gutachter: Prof. Dr. Benno Stein
2. Gutachter: Junior-Prof. Dr. Florian Echtler

Datum der Abgabe: 21. April 2016

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Weimar, 21. April 2016

.....
Sebastian Köpsel

Zusammenfassung

Clickbait sind vorwärts referenzierende und sensationalistische Überschriften, die das Ziel haben den Leser zu ködern. In dieser Arbeit werden verschiedene Ansätze zur automatischen Clickbait-Erkennung von untersucht. Dazu wurde ein Korpus der aus Twitter-Nachrichten mit 2992 Einträgen entworfen, welcher von 3 Annotatoren annotiert wurde. Dabei konnte Clickbait und Nicht-Clickbait mit dem Random Forest Klassifikator mit einer ROC-Area von bis zu 0.826 unterschieden werden. Darauf aufbauend wurde ein Webservice implementiert, der die automatische Bewertung von Tweets über eine Server-API gestattet.

Inhaltsverzeichnis

1	Einleitung	1
2	Verwandte Arbeiten	4
2.1	Nachrichtenüberschriften	4
2.2	Clickbait-Erkennung	5
2.3	Linguistische Forschung über Clickbait	6
2.4	Psychologische Funktionsweise von Clickbait	7
2.5	Listicles	9
3	Clickbait Analyse	10
3.1	Definition Clickbait	10
3.2	Stufen	12
3.3	Listicles	14
3.4	Andere Formen	15
3.5	Meinungen	15
4	Korpuserstellung	17
4.1	Ansätze	17
4.2	Sampling Prozess	18
4.3	Annotationsprozess	21
4.4	Analyse der Annotationen	22
4.5	Clickbait-Verteilung pro Herausgeber	24
4.6	Tweetlänge	26
4.7	Vorwärtsrefrenzierung	26
4.8	Listicles	26
4.9	Andere Formen von Clickbait	27
5	Clickbait-Erkennungsmodell	28
5.1	Attribute	28
5.2	Vorverarbeitung	36
5.3	Attributauswahl	38
5.4	Klassifikatoren	39

5.5	Werte zur Leistungsmessung	41
6	Evaluierung	44
6.1	Hauptexperiment	44
6.2	Vertiefende Experimente	51
6.3	Manuelle Attributmenge	58
6.4	Fehleranalyse	59
7	Webservice	64
7.1	Motivation	64
7.2	Umsetzung	64
8	Zusammenfassung und Ausblick	67
	Literaturverzeichnis	69
A	Appendix	75
A.1	Part of speech labels	75
A.2	X^2 -Verfahren Top 100 Attribute	81
A.3	Selbsterstellte Attributmenge	82

Danksagung

Ich danke den Betreuern meiner Arbeit, Prof. Dr. Matthias Hagen und Dr. Martin Potthast.

Kapitel 1

Einleitung

Nachrichtenüberschriften sind traditionell klare, kurze Zusammenfassungen des Inhalts des dazugehörigen Artikels. Sie haben die Funktion, den Leser über das Kommende zu informieren, so dass er bewusst entscheiden kann, ob er den Inhalt lesen möchte. Gerade in sozialen Medien, wie Twitter oder Facebook, müssen Überschriften gegen eine große Konkurrenz von Themen und Inhalten bestehen. Dort sieht der Nutzer nur Überschriften und Teaser, in Form von Bildern, Videos und Kurzzusammenfassungen der Artikel. Da sich News-Herausgeber über Werbung auf ihren Webseiten finanzieren, sind Besucher ihrer Onlinepräsenz für sie entscheidend. Dies hat zur Entwicklung eines Überschriftenstiles geführt, der die traditionelle Aufgabe einer Überschrift, den Leser zu informieren, nicht mehr erfüllt. Stattdessen versuchen die Überschriften den Leser mit psychologischen Tricks zu Ködern. Dieser Stil wird als *Clickbait* bezeichnet. Clickbait, zu deutsch „Klickköder“ ist im Kontext dieser Arbeit eine Nachrichtenüberschrift mit einem Link, welche das Ziel hat, den Leser zum Verfolgen ihres Verweises zu ködern. Hauptwerkzeug dabei ist ein reißerischer, emotionaler und vorwärts referenzierender Stil.

„This man raised millions for the Conservative party. He’s just been knighted“ An diesem Beispiel für Clickbait lässt sich erkennen, wie auf den Inhalt des Artikels verwiesen wird, ohne genau zu nennen um wen es sich handelt. Dass es auch anders geht zeigt „Trade union leader Kenny knighted“. Diese Überschrift behandelt dasselbe Thema, jedoch wird der Sachverhalt konkret beschrieben.

Die Effektivität von Clickbait steht außer Frage, so konnten Seiten wie Upworthy¹ und BuzzFeed² einen großen Erfolg durch die Verbreitung von Clickbait erreichen. Unsere Recherchen haben jedoch ergeben, dass Clickbait oft als irreführend, sensationalistisch und enttäuschend wahrgenommen wird. Clickbait

¹upworthy.com/

²buzzfeed.com

ist so unbeliebt, dass sich der Twitteraccount @SavedYouAClick, der sich dem Auflösen von Clickbait verschrieben hat, 210.000 Follower erreichen konnte³.

Aus den Erkenntnissen dieser Arbeit entstand bereits eine wissenschaftliche Veröffentlichung, an der Prof. Matthias Hagen und Dr. Martin Potthast mitgewirkt haben[1].

Alle in dieser Arbeit gezeigten Beispiele sind aus einem von uns, für diese Fragestellung, kompiliertes Korpus entnommen. Dieses enthält die gewerteten Tweets, sowie weitere Daten. Die gezeigten Beispiele wurden je nach Kontext von der Mehrheit der Annotatoren als *Clickbait* beziehungsweise *Nicht-Clickbait* gewertet.

Wir möchten Ansätze schaffen, die es ermöglichen, aus der Informationsflut der Sozialen Medien Clickbait zu erkennen und zu filtern. Dies soll die Gelegenheit bieten, Soziale Medien und News-Feeds auf *ehrliche* Überschriften zu reduzieren. Es ist nicht das Ziel dieser Arbeit, „Boulevardjournalismus“ oder „Regenbogenpresse“ zu erkennen. Wir werden in dieser Arbeit zeigen, dass Personen unterschiedliche Schwerpunkte in der Bewertung von Clickbait setzen, was bei der Betrachtung dieses Themas zu beachten ist.

Daraus ergibt sich das Thema dieser Arbeit: wie gut können wir Clickbait, insbesondere auf Twitter, mit Hilfe von Machine Learning erkennen? Dazu werden wir auf folgende Fragen eingehen:

- Lässt sich ein allgemein gültiges Clickbait-Modell erstellen?
- Ist es praktikabel, Clickbait-Erkennung zu personalisieren, also für die Bedürfnisse einer einzelnen Person anzupassen?

Der Tweet ist das Element, das der Leser zunächst zu sehen bekommt. Der Tweet enthält eine Überschrift, diese ist jedoch nicht immer mit der des verlinkten Inhalt identisch, und einen Link. Oft enthält sie auch einen weiteren Aufmerksamkeitserreger in Form eines Bildes, Videos, einer automatisch generierten Seitenvorschau, wie im Falle Twitters, oder anderen Medien. Um den noch nicht wissenschaftlich tiefer behandelten Begriff Clickbait besser zu beschreiben und seine Auswirkungen zu erforschen, haben wir eine Meinungsanalyse anhand von Blogeinträgen zum Thema Clickbait durchgeführt. Weiterhin haben wir das, unseres Wissens nach, erste Korpus erstellt, das speziell für die Erkennung von Clickbait entworfen wurde. Dieses wurde bereits von drei Personen annotiert, wobei jeder Tweet des Korpus auf einer Skala mit 4 Abstufungen zugeordnet wurde, wie stark er versucht, den Leser zu ködern. Ein von uns entworfenes Programm gestattet es uns, durch einige Selektoren Annotationen zu vergleichen, zu erstellen und bei Bedarf zu korrigieren.

³twitter.com/SavedYouAClick Stand 24.02.2016

Diese Arbeit ist wie folgt aufgebaut: Zunächst werden bekannte Ansätze zur Erkennung und Funktionsweise von Clickbait in vorherigen Arbeiten im Kapitel 2 untersucht. Dabei wird auf, unter anderem, auf die psychologische Funktionsweise und die linguistischen Grundlagen von Clickbait eingegangen.

Kapitel 3 wird Clickbait als Phänomen tiefer behandeln. Dazu wird Clickbait differenziert definiert und auf dessen Merkmale eingegangen. Anschließend werden die Unterschiede zwischen Clickbait und traditionellen Nachrichtenüberschriften analysiert. Um die öffentliche Meinung von Clickbait detailliert zu untersuchen, haben wir einige dutzend, Clickbait behandelnde, Blogeinträge verbreiteter Plattformen wie Wordpress, Blogspot und Tumblr untersucht.

Auf das Korpus und dessen Annotation wird in Kapitel 4 genauer eingegangen. Dabei werden die Erstellung, der Annotationsprozess, sowie dessen Struktur und Besonderheiten angesprochen. Das Korpus wurde benutzt, um verschiedene Maschinelle Lernen Experimente durchzuführen. Dazu wurde das *Maschinelle Lernen* Framework Weka⁴ verwendet. Hierbei wurden neben den oben genannten Fragestellungen der Einfluss verschiedener anderer Faktoren untersucht. So gehen wir auf Attributauswahl, Lernkurve bezüglich des Anteils der verwendeten Tweets und den Einfluss des Erstellers eines Tweets ein, um Abschließend eine detaillierte Fehleranalyse durchzuführen. Als Klassifikatoren wurden Naive Bayes, Random Forest, J48⁵ und Logistische Regression verwendet und verglichen. Auf die Details dieser Experimente wird in Kapitel 5 eingegangen. Im folgenden Kapitel 6 wird auf die Ergebnisse dieser Experimente untersucht. Um Grundlagen für eine praktische Anwendung unserer Ergebnisse zu schaffen, wurde auf Grundlage unserer Ergebnisse ein Webservice erstellt, der es auf Basis einer Rest-API ermöglicht, Tweets und beliebige Texte zu klassifizieren. Dessen Aufbau und Funktion werden in Kapitel 7 erläutert. Anschließend werden in Kapitel 8 die Ergebnisse noch einmal zusammengefasst und ein Ausblick auf zukünftige mögliche Aufgaben der Forschung gegeben.

⁴cs.waikato.ac.nz/ml/weka/ Stand: 14.02.2016

⁵Die Implementation des C4.5 Algorithmus in Weka.

Kapitel 2

Verwandte Arbeiten

In diesem Kapitel wird aufgezeigt, welche für diese Arbeit relevanten Erkenntnisse bereits gewonnen wurden. Zunächst wird auf Überschriften eingegangen, um danach in Abschnitt 2.1 die Veränderungen in der Medienlandschaft zu erläutern, die zum Erfolg von Clickbait geführt haben. Anschließend werden vorherige Versuche zur Clickbait-Erkennung untersucht. Daraufhin werden linguistische und psychologische Merkmale analysiert, um zuletzt eine Sonderform von Clickbait, die *Listicles* zu thematisieren.

2.1 Nachrichtenüberschriften

Der Überschriftentext ist der entscheidende Teil einer Botschaft. Mit ihm wird das Interesse geweckt, Medien haben oft nur eine unterstützende Funktion. Hier soll zunächst auf Regeln eingegangen werden, die eine klassische Überschrift erfüllen *sollte*. Diese Regeln stellen in dieser Arbeit den Gegenentwurf zu Clickbait dar. Eine Überschrift, die diese Richtlinien erfüllt, ist im Allgemeinen kein Clickbait. Daniel Dor definierte in seiner Arbeit Regeln für *gute* Überschriften[2]. Diese sind demnach üblicherweise kurze, telegrammartige Zusammenfassungen oder Zitate. Er stellt einige Regeln auf, wie Überschriften gestaltet sein sollte:

- So kurz wie möglich
- Klar, leicht zu verstehen und eindeutig
- Interessant
- Übermittelt neue Information
- Setzt keine dem Leser unbekannte Information voraus
- Enthält Namen und Konzepte mit hohem Nachrichtenwert für den Leser
- Enthält keine Namen und Konzepte mit niedrigem Nachrichtenwert für den Leser

- Verbindet die Geschichte mit älteren Erwartungen und Annahmen
- Verbindet die Überschrift und die Geschichte in einer angemessenen Weise

Dor beschreibt auch den Stil typischer Boulevardüberschriften, da Clickbait jedoch eigenen Regeln folgt, soll an dieser Stelle darauf nicht weiter eingegangen werden.

Nachrichtentypen

In „Doing Well and Doing Good“ zeigt Patterson 2000 eine zunehmende Abwendung der Nachrichtenverlage von Hard News in Richtung Soft News. Soft News wird an Anlehnung an die Arbeit Bloms[3] wie folgt definiert:

[Soft news is] “typically more personality-centered, less time-bound, more practical, and more incident-based than other news” and associated with “sensationalism”, “human-interest” or “news you can use.”

Dabei lässt sich Clickbait als Bestandteil von Soft News verstehen, da es die Merkmale Sensationsmacherei und menschliches Interesse teilt. Clickbait ist jedoch spezifischer, da es sich primär auf den Überschriftenstil bezieht, während sich Soft News auf Nachrichtenthemen bezieht. Der Anstieg des Anteils an veröffentlichten Soft News im Verhältnis zu den traditionelleren Hard News zeigt sich in zunehmender Sensationsmacherei, Aufmerksamkeit für „Human Interest“, Kriminalität und Katastrophen in Nachrichten und weniger Fokus auf Themen öffentlichen Interesses.¹ Patterson konnte auch einen Anstieg von Selbstreferenz der Autoren in News nachweisen, ein für Clickbait typisches Mittel[4].

2.2 Clickbait-Erkennung

In der Vergangenheit gab es bereits private Versuche Clickbait zu erkennen. Die Browser-Plugins *Downworthy*² und *ClickBaiter*³ versuchen mittels regulärer Ausdrücke, typische Clickbait-Ausdrücke und Wendungen auf Webseiten zu erkennen. Downworthy versucht dabei, typische Wendung zu finden und zu ersetzen, um sie lächerlich zu machen. So wird dabei beispielsweise die Phrase

¹Dies bezieht sich in der Arbeit Pattinsons auf das öffentliche Interesse nicht berührende Kriminalität und Katastrophen, sondern sensationalistische Boulevardthemen

²downworthy.snipe.net/ Stand: 13.02.2016

³github.com/georgeportillo/ClickBaiter Stand: 13.02.2016

„You Didn’t Know Exist“ zu „No One Gives a Shit About“. Beide Plugins gehen nicht auf die konkrete Entstehung der von ihnen verwendeten Regeln ein, jedoch ist bei beiden der Quellcode offen. Wir verwenden die von Downworthy verwendeten regulären Ausdrücke und Phrasen als Attribute in unserem Modell. Damit können wir die Leistung dieser Attribute als Basislinie verwenden, da sie einen einfachen Ansatz zur Erkennung von Clickbait darstellen, indem sie typische Redewendungen sammeln und das Vorkommen dieser Wendung immer als Clickbait werten.

Chen et al. unterstreichen, dass eine automatische Clickbait-Erkennung nötig ist, um Leser journalistischer Inhalte vor Fehlinformation und Sensationalismus zu schützen[5]. Parallel zu unserer Arbeit untersuchen Chen et al. theoretisch verschiedene Ansätze zur automatischen Clickbait-Erkennung und kommen dabei zu dem Schluss, dass ein kombinierter Ansatz aus lexikalischer- und semantischer Analyse des Textes, Analyse des Nutzerverhaltens und Bildanalyse am erfolgversprechendsten ist[5].

Facebook hat, weil Clickbait zu häufig in den News Feeds seiner Benutzer auftauchte, Mechanismen eingeführt, um diesen zu filtern[6]. Dabei werden hauptsächlich Attribute benutzt, die uns nicht zur Verfügung stehen. Verwendet wird beispielsweise, wie viel Zeit der Nutzer auf der Zielseite verbringt, bis er zu Facebook zurückkehrt, oder wie viele Besucher des Links diesen dann auch ihrerseits teilen.

2.3 Linguistische Forschung über Clickbait

Eine Studie von Blom und Hansen über die *Clickbaitisierung* dänischer Nachrichten fand 17.2% Clickbait[3] bei der Untersuchung von 100.000 Überschriften von zehn verschiedenen dänischen News-Webseiten. Fokus gelegt wurde dabei auf den vorwärts referenzierende Aspekt von Clickbait. Blom und Hansen verwendeten zur Erkennung von Clickbait hauptsächlich die dänischen Wörter für „Hier“, „Diese“ und „Wir“.⁴ Der größte Anteil von Clickbait in sogenannten „soft content“ Kategorien „Sport“, „Wetter“, „Lifestyle“ und „Gadget“ gefunden wurde. Es wurde von Blom und Hansen festgestellt, dass zunehmend mehr Boulevardzeitungsstil auf den untersuchten Nachrichten Webseiten Verwendung fand, zudem wurde ein Trend zur Kommerzialisierung bemerkt. Damit haben Blom und Hansen die Grundlage für eine automatische Erkennung von Clickbait gebildet, sie erkennen jedoch auch, dass ihre Ergebnisse nicht völlig umfassend sind, sondern eher als Teil einer Tendenzbewertung verstanden werden sollten. Das von uns kompilierte Korpus enthält jedoch, jedenfalls nach

⁴her, sådan, derfor, så and dette, denne, dette

unserer Definition, mehr Clickbait, da wir uns nicht auf den vorwärts referenzierenden Stil beschränkt haben.

2.4 Psychologische Funktionsweise von Clickbait

Elly Ifantidou konnte zeigen, dass Leser Überschriften, die nicht informativ, dafür jedoch kreativ sind, bevorzugen[7]. Dazu wurden 137 Leserreaktionen auf 1310 ausgewählte englische Überschriften amerikanischer und britischer Herausgeber, darunter *The New York Times*, *The Daily Telegraph* und *The Observer* ausgewertet. Dabei zeigte sich, dass viele Leser Länge, Klarheit und Informationsgehalt nicht als wichtige Attribute einer Überschrift ansehen.

Bei der Untersuchung, welche Faktoren für Viralität von Online-Inhalten entscheidend sind, sind Berger et al. zu dem Schluss gekommen, dass besorgnis- und wuterregende Geschichten häufiger in Mailinglisten geteilt wurden[8]. Dass solche emotionale Geschichten und Überschriften mehr Aufmerksamkeit erfahren und dementsprechend Menschen besser ködern, konnte auch von Guerini et al. gezeigt werden[9]. Dies resultiert auch in höherer Viralität, also dem „epidemischen“ Ausbreiten eines Artikels, Videos, Bildes oder Ähnlichem.

Da Clickbait ebenfalls oft an Emotionen appelliert, lässt sich schließen, dass Emotionen eingesetzt werden um Menschen stärker anzusprechen als es neutralere Überschriften vermögen, und sie somit zum Lesen des Artikels zu ködern, aber auch um die Chancen zu erhöhen, dass die Botschaft auf sozialen Netzwerken geteilt wird[8, 9].

Um Emotionen abzubilden wird von Guerini et al. das VAD Modell verwendet[9]. Das VAD Modell benutzt dazu die Faktoren Anziehungskraft, Erregung und Dominanz[10]. Es wurde von Guerini gezeigt, dass Viralität oftmals bei bestimmten Konfigurationen dieser Werte entstand[9].

Warriner et al. haben für 13,915 englische Worte Werte für die VAD-Attribute ermittelt[11]. Wir verwenden diese Werte, um die Botschaften auf der VAD-Skala einzuordnen.

In einer Untersuchung über die News-Präferenzen von verschiedenen Lesergruppen konnten Hensinger et al. zeigen werden, dass emotional aufgeladene Inhalte beliebter bei Lesern sind[12]. Besonders genannt werden dabei Entertainment und Kriminalität.

Bei der Verwendung von Emotionen in der automatischen Clickbait-Erkennung sind verschiedene Aspekte zu beachten. So konnte Kouloumpis et al. zeigen, dass Stimmungserkennung mit den Part-of-speech Taggern auf üblichen Twiternachrichten nicht gut funktioniert[13]. Da Clickbait jedoch eine spezielle Form der Überschrift darstellt, haben wir das Framework Stanford CoreNLP Framework, welches ebenfalls Funktionen zur Stimmungserkennung be-

reitestellt, verwendet[14]. Dies ermöglicht uns zusätzlich zum VAD-Modell die Stimmung von Botschaften in unserem Erkennungsmodell abzubilden.

Es lassen sich jedoch verschiedene Probleme vorhersagen[1]:

- Nachrichtenüberschriften und Clickbait geben meist nicht die Meinung des Autors wieder, eine Voraussetzung, von welcher in der Stimmungsanalyse ausgegangen wird
- Clickbait verwendet nicht immer stimmungserzeugende Begriffe
- Boulevardzeitschriften verwenden auch emotionale Begriffe, veröffentlichen aber nicht zwangsläufig Clickbait

Ein weiterer Erklärungsversuch für Clickbait ist die sogenannte *Fear of missing out*⁵ (FoMO). FoMO bezeichnet die, für betroffene Personen, allgegenwärtige Befürchtung, dass Andere ein gutes Erlebnis haben, bei dem man selbst nicht anwesend ist. Dies äußert sich in dem Verlangen, mit den Aktionen Anderer ständigen Anschluss zu haben[15]. Nadya Khoja vermutet, dass sich Clickbait-Verbreiter dieses Mechanismus bedienen, um Menschen das Gefühl zu geben, dass es sozial notwendig ist, diese Artikel zu teilen[16]. Einige Themen, wie Essen oder Selbstwahrnehmung eignen sich demnach besonders gut um FoMO hervorzurufen. Durch den sensationalistischen Charakter von Clickbait hätten Menschen das Gefühl etwas zu verpassen, falls sie nicht auf den verlinkten Inhalt der Botschaft folgen.

In journalistischen Artikeln wird häufiger die „Curiosity Gap“ als entscheidender Faktor in der Funktionsweise von Clickbait genannt[17]. Dieser Ausdruck bezieht sich auf die Arbeit Loewensteins[18], wonach Menschen oftmals das Verlangen empfinden, eine entstehende Wissenslücke zu schließen, falls sie auf eine solche treffen. Bei Clickbait trifft dies zu, da ein Thema in der Überschrift angedeutet wird, jedoch erst im Artikel konkretisiert wird.

Davalos et al. sind der Frage nachgegangen inwiefern sich die Titelblattüberschriften von Magazinen, besonders an Frauen gerichtete, in den Jahren 1976 bis 1996 verändert haben[19]. In Abbildung 2.1 werden von Davalos et al. genannte Überschriften abgebildet. Es zeigt sich, dass bereits 1976 Überschriften verwendet wurden, die heutzutage als Clickbait bezeichnet werden würden. Clickbait ist also keine neue Erfindung, sondern findet schon seit mindestens 40 Jahren Verwendung, um Menschen zum Lesen bestimmter Artikel oder zum Kauf von Magazinen zu verleiten.

⁵Angst, etwas zu verpassen

Abbildung 2.1: Überschriften aus Frauenmagazinen 1976

- „Passive Men: Love Them or Leave Them or What?“
- „90 Greatest Looks Going: Super Accessories“
- „Latest News on Cosmetic Surgery“
- „How to Kick the Worry Habbit“

2.5 Listicles

Listicles⁶ sind Artikel in Listenform und eine häufig auftretende Form von Clickbait. Weil Listicles oft unter geringen Aufwand erzeugt werden, wird Listicle meist als abwertender Begriff gebraucht[20]. Da Listicles jedoch trotzdem erfolgreich und beliebt sind, analysiert Vijgen anhand von 720 Beispielen, welche Faktoren Listicles erfolgreich machen[21]. Er kommt dabei zu dem Schluss, dass der Erfolg nicht nur am Format selbst liegt, sondern insbesondere an den Faktoren: Suggestion von Autorität, das Maß an Verführung und die Einfachheit der Wahrnehmung, öffentliche Reichweite, Originalität und Diversität. Vijgen stellt weiterhin fest, dass alle von ihm untersuchten Listicles eine Kardinalzahl in der Überschrift enthalten, 85% der Listicles beginnen mit dieser Zahl.

In einem Artikel des *New Yorker* argumentiert Konnikova mithilfe von Messner und Wänker, dass Menschen sich dadurch, dass die zum Lesen benötigte Zeit aufgrund der Nummerierung gut abschätzbar ist, eher darauf einlassen, dies Artikel zu lesen[22, 23]. Des Weiteren stechen Zahlen aus dem Text eher heraus und erregen so optisch Aufmerksamkeit. Als verstärkender Faktor kommt ihrer Einschätzung nach hinzu, dass durch das erfolgreiche „Abschließen“ des Artikels ein Belohnungseffekt entsteht. Dieser Belohnungsfaktor führt wiederum dazu, dass Listen und Listicles als lohnenswert zu lesen empfunden werden. Weiterhin führt sie an, dass die Schätzbarkeit der benötigten Zeit und durch die vorgegebene Kategorisierung der zu erwartenden Inhalte für Menschen sehr interessant ist, da diese Listen „müheles erworbene Daten“ darstellen. Dies führt dazu, dass Listicles oft als etwas Positives und Interessantes wahrgenommen werden und dazu, dass Listicles ein häufig auftretender Artikeltyp sind.

⁶Ein englisches Mischwort aus Liste und Artikel.

Kapitel 3

Clickbait Analyse

Nachdem in dem vorherigen Kapitel auf linguistische und psychologische Mechanismen von Clickbait eingegangen wurde, wollen wir in diesem Kapitel zeigen, in welchen Merkmalen von Clickbait sich dies äußert. Dazu betrachten wir zunächst die Definition von Clickbait, um daraufhin die von uns gewählten Annotationsstufen zu beschreiben.

3.1 Definition Clickbait

Eine genaue, allumfassende Definition von Clickbait ist schwer, vor allem weil dieser Begriff sehr neu ist. Laut „Google Trends“¹ beginnt Clickbait erst 2013 im Internet als Wort zu erscheinen. Dies deckt sich mit unserer Erfahrung bei der Meinungsanalyse, wonach der Begriff Clickbait erst ab 2014 in Blogs und journalistischen Artikeln auftaucht.

Merriam Webster², ein bekanntes englisches Wörterbuch, hat den Begriff Clickbait im Mai 2015 als Wort aufgenommen[24]. Clickbait wird demnach definiert als

„something (such as a headline) designed to make readers want to click on a hyperlink especially when the link leads to content of dubious value or interest“

Für diese Arbeit liegt jedoch nicht der Inhalt des verlinkten Artikels als Kriterium für Clickbait zugrunde. In dieser Arbeit wird Clickbait definiert als: *Tweets die sich eines Tricks bedienen, um den Leser zu einem Klick zu bringen.*

Wikipedias Definition ist unserer Meinung nach zutreffender:

¹google.de/trends/explore?q=clickbait Stand: 13.03.2016

²merriam-webster.com/

„Clickbait is a pejorative term describing web content that is aimed at generating online advertising revenue, especially at the expense of quality or accuracy, relying on sensationalist headlines to attract click-throughs and to encourage forwarding of the material over online social networks. Clickbait headlines typically aim to exploit the ‚curiosity gap‘, providing just enough information to make the reader curious, but not enough to satisfy their curiosity without clicking through to the linked content.“[25]

Das wichtigste Kriterium für Clickbait ist die schon erwähnte Vorwärtsreferenzierung, wie auch schon zuvor von Blom und Hansen festgestellt wurde[3]. Jedoch haben wir noch andere Merkmale gefunden welche das Ziel haben den Leser ködern. Um einen Tweet als Clickbait auszuzeichnen, ist es nicht nötig, dass alle im Folgenden angesprochenen Merkmale auftreten. Vielmehr führt das Zusammenspiel dieser Merkmale dazu, dass ein Tweet als Köder empfunden wird. Dabei ließ sich in unserer Annotationsphase beobachten, dass Annotatoren einzelne Faktoren sehr unterschiedlich gewichten und dementsprechend unterschiedlich einschätzen, ob ein Tweet Clickbait ist, oder vielmehr eine kreative Überschrift. Im Folgenden werden wir verschiedene Merkmale nennen, die für Clickbait typisch sind.

Das neben der Vorwärtsreferenzierung wichtigste Clickbait-Merkmal ist das Nichtinformieren des Lesers über den eigentlichen Inhalt des Artikels. Diese beiden Kriterien werden oft zusammen verwendet. Dabei wird der eigentliche Inhalt des verlinkten Artikels nicht vermittelt, wie in folgendem Beispiel:

Why Shirley Inspired is the one album you should hear this week—video

Die Erklärung, warum *Shirley Inspired* das eine Album der Woche ist, findet sich nur im verlinkten Video. Ebenso lässt sich erkennen, wie die bereits erwähnte *Fear of missing out* verwendet wird. Der Autor appelliert direkt daran, aktuelle Entwicklungen mitzuverfolgen. Es wird angedeutet, dass der Leser den Anschluss verlieren würde, falls dieser das referenzierte Video nicht ansieht. In diesem Tweet wird durch die Begriffe *should* und *this week* eine für Clickbait typische Dringlichkeit erzeugt. Viele Clickbait-Überschriften, wie auch im vorherigen Beispiel, sprechen den Leser direkt an oder beziehen den Autor des Artikels mit ein:

How successful are you at making the most of your time?

An diesem Beispiel ist eine weitere häufige Eigenschaft von Clickbait enthalten: die Frage als implizite Weiterleitung. Es wird in der Überschrift eine Frage gestellt, anstatt eine Information zu liefern. Um die Antwort auf diese Neugier erzeugende Frage zu erhalten, muss der Leser sich die zugehörige Website ansehen.

Clickbait enthält oftmals eine emotionale Komponente. Es wird versucht, den Leser emotional zu involvieren:

Sometimes the most powerful thing you can do is say an unequivocal „no.“

Die Aussage des Tweets ist klar, aber durch die starke emotionale Anziehung in „powerful“, die Autorität des Satzes und das direkte Ansprechen des Lesers wird der Leser geködert. Typischerweise fehlt der Kontext der Aussage, der Inhalt des Artikels ist nicht erkennbar.

Ein ebenfalls häufiger Aspekt ist die Verwendung von Autorität.

You have to see these two babies dancing To Beyoncé’s “7/11”

Tweets wie diese gehen über einen einfachen Hinweis auf die Vorgabe hinaus. Es wird direkter Druck auf den Leser ausgeübt, den dazugehörigen Artikel zu lesen. Der Leser wird unverhohlen aufgefordert, dem Link zu folgen. Wie schon in Abschnitt 2.4 angesprochen, konnte gezeigt werden, dass Dominanz in Tweets häufig mit Viralität in Verbindung steht[9]. Dominanz wird über aktive Verben[26], oft in der Befehlsform, erzeugt.

See the best cities for freelancers in 2015

Dies wird auch als „Call to action“ [26–28] bezeichnet. Call to action bezeichnet das Auffordern des Kunden (im Journalismus der Leser) zu einer Handlung. Diese ist im Fall Clickbaits das Besuchen der Webpräsenz des Erstellers. Während der Leser bei herkömmlichen Überschriften selbst entscheidet, ob ihn das Thema des Artikels interessiert, wird bei Tweets mit einem Call to action versucht Autorität über den Leser auszuüben und ihm diese Entscheidung vorzugeben. Daher wird Call to action als Clickbait-Merkmal betrachtet.

3.2 Stufen

Zu Beginn der Arbeit an unserer Arbeit an Clickbait-Erkennung haben wir ein Pilotkorporus kompiliert und annotiert, bei welchem wir die Annotation auf die zwei Klassen „Clickbait“ und „Nicht-Clickbait“ beschränkt haben. Dieses Pilotkorporus enthielt etwa 1,500 Tweets. Bei der Arbeit mit diesem Pilotkorporus hat sich herausgestellt, dass Clickbait in vielfältigen Formen und Stärken vorkommt.³ Wir haben uns daraufhin dazu entschieden, Clickbait-Stärke in vier Stufen darzustellen, um eine differenzierte Bewertung der Tweets zu erhalten:

Nicht-, Schwacher-, Mittlerer-, und Starker-Clickbait.

³Der Pilotkorporus wurde für diese Arbeit sonst nicht verwendet.

Aus dieser Bewertung lassen sich weitere, für den Zweck der jeweiligen Erkennung anpassbare, Gruppen bilden. Denkbar wäre eine Art von Ampel, welche Tweets in Rot (Starker-CB), Gelb (Schwacher- und Mittlerer-CB) und Grün (Nicht-CB) einordnet. Benutzer einer Clickbait-Erkennung hätten damit die Möglichkeit, auf einen Blick zu erfassen ob Clickbait vorliegt. Ein ähnliches System wurde für den in Kapitel 7 beschriebenen Webservice implementiert. Wir verwenden die balancierte Binarisierung, bei welchem eine Trennung in Clickbait (Medium, Stark) und Nicht-Clickbait (Kein, Schwach) erfolgt.

Anhand der in der Pilotphase gewonnen Erkenntnisse wurden vor der individuellen Annotation des Korpus Kriterien für die die jeweiligen Stufen eingeführt, diese werden im Folgenden beschrieben. Alle der hier als Beispiele verwendeten Tweets enthalten keine Medien und wurde von allen Annotatoren in der entsprechenden Kategorie eingestuft. Auf die für diese Betrachtungen irrelevanten Links wurde verzichtet.

Als Nicht-Clickbait wurden Tweets eingestuft, welche keine der für Clickbait typischen Attribute aufweisen. Meist sind dies klassische Nachrichtenüberschriften, die klar und ohne Doppeldeutigkeit den Inhalt des verlinkten Artikels zusammenfassen und den in Abschnitt 2.1 genannten Regeln entsprechen.

- John Biggs elected as Labour mayor of Tower Hamlets
- Swiss voters overwhelmingly reject federal inheritance tax initiative
- State Dept. says US officials will not meet with Muslim Brotherhood delegation

Die Themen der verlinkten Artikel sind hierbei klar und sachlich genannt. Es wird nicht versucht über einen reißerischen Text Leser zu ködern.

Als schwacher Clickbait wurden dabei insbesondere Tweets gewertet, welche zwar das Thema neutral umreißen, dabei jedoch nicht endgültig beschreiben oder Lücken erzeugen. Einige Tweets erhielten die Einstufung aufgrund ihres reißerischen Tones.

- Amazing messaging tool Slack kills 80 to 100 percent of work emails
- Abortions are way down. Democrats and Republicans should stop congratulating themselves.
- Billionaire Elon Musk can't deny that his companies are getting a ton of help from the government

Bei dem ersten Tweet ist noch erkennbar, dass das Thema des verlinkten Artikels Slack ist, wobei dies ein Nachrichtenwerkzeug ist. Jedoch erhält der Tweet

aufgrund des Fehlens von Informationen, wie die Arbeitsmails reduziert werden und durch den sensationalistischen Character von *amazing* und *kills* erhält dieser Tweet die Einstufung Schwacher-Clickbait.

Als Mittlerer-Clickbait wurden Tweets eingestuft, falls sie das Thema nur noch grob umreißen, jedoch noch Hinweise auf den Inhalt enthalten:

- Some snack bars make steak look cheap
- 15 free online resources that will improve your productivity
- This school had a cute way of telling students they can wear uniforms for genders they identify with

So lässt sich im ersten Tweet erkennen, dass das Thema des Artikels Essen, beziehungsweise Snackbars sind, jedoch nicht viel mehr.

Ein Tweet wurde als „Stark“ eingestuft, falls er absichtlich und reißerisch Informationen vorenthält, gar keinen Kontext enthält oder von den Annotatoren aus anderen Gründen als Stark täuschend wahrgenommen wurde:

- Someone make @taylorswift13 find this
- Here comes (almost) free money
- There's one thing Wall Street interns care about more than which bank they get into

Es lässt sich erkennen, dass im ersten Tweet das Thema des Links nicht genannt wird und mit „this“ eine direkte Vorwärtsreferenzierung stattfindet.

3.3 Listicles

Listicles sind sehr oft Clickbait und werden diesem oft zugeordnet[26, 29]. Wir haben Listicles nicht grundsätzlich als Clickbait gewertet, sondern sie, wie alle anderen Tweets, aufgrund ihrer Köderwirkung bewertet. Die Anzahl der Listienpunkte in dem Artikel ist meist in der Überschrift, in unserem Fall der Text des Tweets, enthalten. Ein Beispiel für ein typisches Clickbait-Listicle wäre „16 perfect Japanese words you need in your life“. Listicles bestehen normalerweise aus einem in Stichpunkte untergliederten Text oder einer Liste von Bildern. Um Seitenaufrufe und somit Werbeeinnahmen zu erhöhen sind, insbesondere bei Bild-Listicles, die Punkte auf mehrere Unterseiten verteilt[29].

3.4 Andere Formen

Neben Listicles gibt es noch andere Formen von Clickbait-Inhalten. Dazu gehören insbesondere Quiz und Videos. Clickbait-Tweets zu Videos enthalten häufig die Aufforderung diese zu schauen, wie bei: „Here’s Ron Burgundy randomly calling a home run by the @Reds’ Ivan De Jesus Jr. Because hey, why not?“. Die Videos sind meist auf den Seiten der Herausgeber eingebettet. Quiz sind weitere Inhalte, die häufig Clickbait sind. Ein Beispiel für eine Clickbait-Überschrift für ein Quiz ist: „Quiz: have you been using emoji all wrong?“.

3.5 Meinungen

Um eine Übersicht über die öffentliche Meinung über Clickbait zu erhalten, haben wir mehrere dutzend Einträge verschiedener bekannter Blog-Plattformen, darunter Wordpress, Blogspot und Tumblr, analysiert .

Der überwiegende Ton ist dabei sehr negativ, es konnten kaum Clickbait-Befürworter gefunden werden. Dabei sind die Hauptkritikpunkte, dass Clickbait als irreführend, sensationalistisch und enttäuschend wahrgenommen wird.

Als problematisch eingestuft wird, dass oftmals komplexe Themen auf aggressive, einseitige Titel beschränkt werden[30–32]. Von vielen Bloggern wird bemängelt, dass Clickbait auch aus einer journalistischen und geschäftspolitischen Sicht schlecht ist. Kara Dennison sagt dazu:

„ It’s poor writing. It’s poor business sense. It’s poor marketing.
And it’s getting us accustomed to it[..]“[33]

Dennison argumentiert, dass journalistisch unzureichende Inhalte langfristig Leser abschrecken und das treue Leser die nicht durch sensationalistische Artikel geködert werden besser für einen Herausgeber sind. Mit der Meinung, dass Clickbait schlecht für das Geschäft ist, steht sie nicht allein[34, 35]. In einem Punkt stimmen viele Blogger überein: Clickbait funktioniert[33, 36, 37]. Dies wird auch von einigen Journalisten bestätigt, die mit A/B Tests verschiedene Überschriften testen und mit Clickbait die besten Besucherzahlen erzielen[38, 39]. Deshalb gibt auch die Einstellung, dass der Inhalt des zum Tweets gehörenden Artikels entscheidend ist, ob eine Überschrift Clickbait ist. Da Clickbait gut funktioniert, werden oft inhaltlich und stilistisch gute Artikel mit Clickbait-Überschriften beworben, um mehr Besucher zu erreichen[40]. Es gibt jedoch auch Stimmen, die Clickbait verteidigen, mit dem Argument, dass Clickbait für Herausgeber notwendig ist, um sich erhalten zu können[36]. Auch das Clickbait dazu verwendet werden kann, erst das Interesse am Thema zu wecken, um daraufhin einen tiefer gehenden Artikel zu bieten wird angeführt[36].

Zudem eignet sich Clickbait tendenziell auch um politische, oder anderweitig von Lesern nicht präferierte, Inhalte zu übermitteln, da sie zum Lesen des Artikels geködert werden[36]. Als Argument für Clickbait wird auch angeführt, dass die Eigenschaft von Clickbait, Argumente und Thesen kurz und einprägsam zusammenzufassen, und im Falle von Listicles auf eine überschaubare Anzahl klarer Unterpunkte zu reduzieren, als pädagogische Gelegenheit verstanden sollte[41].

Aufgrund des schlechten Rufes von Clickbait glauben wir, dass es durchaus das Bedürfnis gibt Clickbait aus dem allgemeinen Nachrichtenstrom filtern zu können. Sensationalistische Überschriften, die psychologische Tricks verwenden um den Leser zu ködern, sollten filterbar sein, ähnlich wie gleichermaßen unerwünschter Spam.

Kapitel 4

Korpuserstellung

Im Folgenden gehen wir auf das Korpus ein. Dabei werden mögliche Ansätze zur Erstellung eines Korpus der für Clickbait-Erkennung geeignet ist geschildert. Weiterhin wird beschrieben wie unseres Korpus erstellt wurde und welche Daten es enthält. Daraufhin wird der Annotationsprozess erläutert und auf die Urteilerübereinstimmung eingegangen.

4.1 Ansätze

Bei der Planung des Korpus sollten einige Bedingungen erfüllt werden, die sicherstellen, dass das entstehende Clickbait-Modell möglichst umfassend, praktisch anwendbar und Allgemein gültig ist. Es ist natürlich notwendig, dass sowohl Clickbait- als auch Nicht-Clickbait-Überschriften von verschiedenen News-Herausgebern gesammelt werden, um maschinelles Lernen zu ermöglichen. Während auch andere Online- Herausgeber, zum Beispiel Blogger, Clickbait verwenden um ihre Leserzahlen zu erhöhen, wurden News-Herausgeber gewählt, weil diese am häufig und regelmäßig Clickbait verbreiten. Ein hoher Bekanntheitsgrad der Quelle ist vorteilhaft, um ein Modell zu erzeugen, welches bei Verwendung viele Menschen erreichen kann. Es ist wichtig, mehrere Herausgeber als Quellen für Tweets zu verwenden, um eine Voreingenommenheit gegenüber dem Stil von Herausgebern oder Autoren zu vermeiden. Es bieten sich also News-Aggregatoren an, da bei diesen mehrere Herausgeber in der selben Form dargestellt werden. Es ist vorteilhaft, wenn die von den Herausgebern verfassten Überschriften verwendet werden oder dass die Herausgeber selbst ihre Einträge öffentlich machen können. Dieser Punkt ist notwendig, um sicherzustellen, dass ein Artikel nicht von Drittparteien als Clickbait eingestellt wird, wie es etwa auf Reddit möglich ist, da dort beliebige Personen einen Link unter einer selbstgewählten Überschrift verbreiten können. Weiterhin ist dieser Schritt erforderlich um das Einwirken Dritter zu vermeiden. Um

eine Themenspezialisierung zu vermeiden wurden spezialisierte Aggregatoren, zum Beispiel Tech-News-Seiten, nicht in Betracht gezogen. Auch der Faktor der sozialen Netze ist wichtig. Wie zuvor dargestellt, ist Viralität ein wichtiger Bestandteil von Clickbait, der nur gegeben ist, wenn direkt beim Leser die Möglichkeit besteht, die Inhalte auf den einschlägigen Netzwerken zu teilen. Es kamen somit primär folgende mögliche Quellen in Frage: Facebook und Twitter. Andere mögliche Quellen (unter vielen weiteren), welche die genannten Kriterien jedoch nicht erfüllen, sind RSS Feeds (keine sozialen Funktionen und geringe Verbraucherverbreitung), Google News (Auswahl der gezeigten Artikel durch Google), Reddit (Dritteinstellung und Internet-Trolle). Wir haben uns für Twitter entschieden, da es mit rund 320 Millionen monatlichen Nutzern weit verbreitet ist und weil alle großen News-Herausgeber hier aktiv Inhalte publizieren[42]. Zudem stellt Twitter eine gute API bereit, welches die Korpuserstellung vereinfacht. Facebook versucht aktiv, Clickbait aus den News-Feeds seiner Nutzer zu filtern und ist daher nicht als Quelle geeignet[6]. Um die Aufgabe besser einschätzen zu können und um Clickbait besser kennenzulernen, wurde am Anfang des Projektes ein erstes Pilotkorpus mit etwa 1500 Einträgen erstellt. Dieser wurde unter Verwendung einiger bekannter News-Seiten, zum Beispiel BBC, Reuters und CNN erstellt sowie einige Herausgeber, welche für Clickbait bekannt sind, unter anderem BuzzFeed. Mithilfe dieses Korpus konnten wir Kriterien für Clickbait diskutieren und erste Attribute testen.

4.2 Sampling Prozess

Das für diese Arbeit verwendete Korpus wurde in einem zweiten Schritt erstellt. Um für die Clickbait-Erkennung relevante, also keine Konversations-Tweets und vielgelesene Tweets untersuchen zu können und eine breite Auswahl an Überschriftenstilen und Themen zu erhalten, haben wir eine Liste der 20 meistgelesenen und geteilten News-Herausgeber auf Twitter im Jahr 2014 gebildet. Dafür wurde die von newswhip.com veröffentlichten Daten bezüglich der Beliebtheit und Verbreitung von Twitter-Nutzern ausgewertet. Newswhip.com ist eine Webseite, die monatliche Daten der Social Media Plattformen Facebook, Twitter und LinkedIn gesammelt und ausgewertet hat. Die ausgewählten Publisher sind nach Anzahl der Retweets¹ der von ihnen veröffentlichten Tweets in unserem Korpus in Tabelle 4.1 dargestellt. Bei leeren Feldern in der Abbildung 4.1 stand uns die Anzahl der Retweets für diesen Monat nicht zur Verfügung. Durch die Auswahl an bekannten Herausgebern konnten wir sicherstellen, dass wir vielgelesene Inhalte erhalten. Mithilfe der Twitter-API²

¹Retweets sind eine Teilungsmöglichkeit auf Twitter, bei der Tweets von einem anderen Nutzer unverändert von dem jeweiligen Nutzer nochmal getweetet wird.

Tabelle 4.1: Top 20 Twitter Publisher nach Retweets in Tausend

Herausgeber	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dec	Total
ABC News	791	1149	1759	1559	1624	1803	1907	1678	1668	1365	1279	1060	17641
BBC News	3	2481	2715	2866	3124	3245	3557	3392	3790	4020	3960	3752	39606
Bleacher Report	582	566	682	725	741	843	843	827	1088	1117	1179	1036	10228
Business Insider	458	502	537	492	470	495	545	520	595	711	564	630	6519
BuzzFeed	756	800	812	839	738	740	993	850	857	897	861	820	9962
CNN	1179	1155	1207	1247	1245	1169	1385	1307	1352	1532	1274	966	15019
Daily Mail Online	510	477	512	536		652	733	670	752	688	786	627	6942
ESPN	441	399	502	523	531	488	579	534	759	782	690	666	6894
Forbes	798	813	962	893	958	1028	1112	1129	1176	1102	778	752	11501
Fox News	587	540	681	824	831	921	948	961	967	1036	993	874	10164
Huffington Post	959	721	854	771	834	933	1142	1137	1123	1170	1022	900	11566
Mashable	1421	1460	1656	1829	1768	1834	1952	1657	1857	1816	1776	1600	20624
NBC News	605	640	779	751	718	746	860	895	954	953	925	874	9698
The New York Times		1281	1836	1997	2113	2285	2477	2355	2420	2598	2222	2179	23763
The Telegraph		502	593				889	890	895	937	909	821	6435
The Independent	339			392	461	445	693	672	598	746	821	671	5838
The Guardian		1067	1157	1130	1226	1324	1420	1266	1370	1511	1351	1189	14009
Washington Post		387	400	410	417	507	620	734	746	801	701	644	6366
Wall Street Journal		559	643	630	663	684	809	773	878	904			6543
Yahoo		568	663	587	669	767	899	1002	838	818	773	575	8160

wurden die Tweets der ausgewählten Herausgeber aus dem Zeitraum der Woche vom 08. Juni bis zum 14. Juni 2015 gesammelt. Dabei haben wir jedoch nur Tweets gespeichert, welche einen Link enthalten, da ein Tweet ohne Link kein Clickbait sein kann. Da die Herausgeber auch mit Lesern über Twitter kommunizieren und dabei typischerweise keine Links verwenden, hilft uns diese Einschränkung, um für uns unwichtige Tweets zu entfernen. Um derartige Konversations-Tweets auszuschließen wurden weiterhin nur Tweets für das Korpus ausgewählt die keine Antwort auf einen anderen Tweet sind.

Die Anzahl an publizierten Tweets unserer Herausgebersauswahl in der untersuchten Woche ist in Tabelle 4.2 zu sehen. Um die Anzahl der Tweets zu beschränken, haben wir für unser Korpus aus dieser Menge 150 Tweets pro Herausgeber zufällig ausgewählt. ESPN hat in der ausgewählten Woche nur 142 Tweets mit Link geteilt. Daher hat unser Korpus eine Gesamtgröße von 2992 Tweets. Zur Kommunikation mit der Twitter-API verwendeten wir die Java-Bibliothek „twitter4j“. Die Tweets werden von der Twitter API als JSON-Objekte repräsentiert und wurden auch in dieser Form gespeichert. Die JSON-Objekte enthalten eine Vielzahl an Attributen, die den Tweet, den Ersteller des Tweets und die enthaltenen Medien beschreiben. Dabei wurden nicht alle enthaltenen Attribute in dieser Untersuchung verwendet, da nicht davon auszugehen ist, dass beispielsweise das Profilbild des Erstellers eines Tweets Auswirkungen auf die Clickbait-Erkennung hat. Falls der Tweet ein Bild enthalten hat, haben wir dieses gespeichert. Damit gewährleisten wir die Vollständigkeit

²dev.twitter.com/rest/public

Tabelle 4.2: Gesammelte Tweets 08. Juni - 14. Juni 2015 der ausgewählten Herausgeber

Herausgeber	Tweets	Herausgeber	Tweets
The New York Times	875	Washington Post	691
Mashable	803	The Independent	530
Business Insider	779	Daily Mail Online	516
Huffington Post	770	NBC News	408
Wall Street Journal	747	Fox News	378
The Guardian	744	CNN	345
Forbes	721	ABC News	279
The Telegraph	699	Bleacher Report	196
BuzzFeed	695	Yahoo	195
BBC News (UK)	694	ESPN	142

des Korpus, auch wenn das Bild online gelöscht wird. Andere Medien, etwa Videos wurden in unserem Korpus nicht von den Herausgebern verwendet. Auf andere Medien die durch Twitter automatisch eingebunden werden wurde keine Rücksicht genommen.

Wir haben den Internetservice „Imagga“³ verwendet um die Bilder zu annotieren. Imagga ermöglicht es, Bilder über eine Web-API durch ein Maschinelles Lernen-Modell zu taggen. Funktionsweise und Attribute des Imagga-Modells werden von dessen Anbietern nicht näher erläutert. Die erhaltenen Tags beschreiben den Inhalt des Bildes. In Abschnitt 5.1 Abbildung 5.3 ist ein Beispiel dargestellt. Diese Informationen konnten wir in unserem Modell verwenden, um den Inhalt der Bilder darzustellen. Von Imagga wurden 3551 einzigartige Bildtags gefunden. Die 5 Häufigsten sind in Abbildung 4.3 dargestellt.

Des Weiteren wurde der Inhalt der gelinkten Internetseite gespeichert. Durch das Auswerten einer einzelnen Woche treten in dem Korpus einige Themen häufig auf, da sie in dieser Woche die Nachrichten geprägt haben. Zu diesen Themen gehört die Suche nach entflohenen Häftlingen (26 Tweets), die Frage ob eine Frau kaukasischer oder afrikanischer Abstammung ist (42 Tweets) und Basketball (123 Tweets). In 59 Tweets wird der Basketballspieler LeBron und in 62 Tweets Apple (meist der Konzern) erwähnt.

Tabelle 4.3: Die 5 häufigsten Bildtags

- people
- person
- happy
- man
- adult

³imagga.com/

Abbildung 4.1: Annotationsprogramm

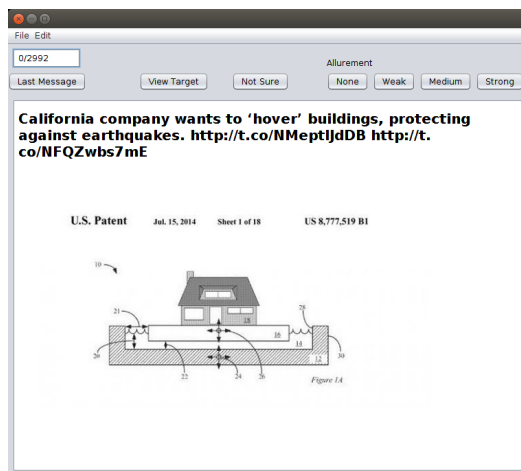
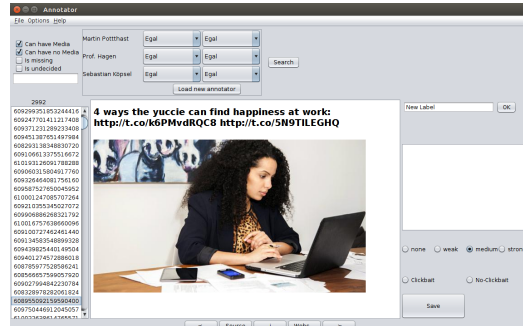


Abbildung 4.2: Annotationsprogramm
2



4.3 Annotationsprozess

Das Korpus wurde von drei Annotatoren annotiert. Bei den Annotatoren handelt es sich um einen Professor der Fakultät, einen wissenschaftlichen Mitarbeiter und den Autor dieser Arbeit. Dabei wurden die Tweets von jedem Annotator in eine der vier in Abschnitt 3.2 beschriebenen Stufen zugeordnet.

Die Annotationsphase

Die Annotation wurde mithilfe eines selbst geschriebenen Annotationsprogrammes bewerkstelligt, welches in Abbildung 4.1 zu sehen ist.

Die Tweets des Korpus werden in zufälliger und achronologischer Reihenfolge gezeigt. Es wird jeweils ein Tweet in einer einfachen Darstellung angezeigt, wobei der Text und, falls vorhanden, das Bild gezeigt werden. Falls der Tweet Medien enthält, welche keine Bilder oder Ähnliches sind, werden diese nicht angezeigt, es wird jedoch auf die Existenz dieser Medien hingewiesen. Da jedoch keine Medien außer Bildern von den Tweets verwendet wurden, fand diese Funktion keine Anwendung. Der gesamte Tweet-Text wird angezeigt, einschließlich der enthaltenen Links. Weitere Informationen, wie Ersteller oder Zeit der Erstellung, werden nicht angezeigt, um eine Voreingenommenheit der Annotatoren zu vermeiden. Dem Annotator konnte zwischen den vier von uns festgelegten Stufen der Clickbait-Stärke, Nicht-, Schwacher-, Mittlerer- und Starker-Clickbait entscheiden. Das Annotationsprogramm wurde im späteren Verlauf weiter entwickelt, um Tweets anhand verschiedenen Parametern zu suchen, Tweets mit zusätzlichen Labels auszustatten (bspw. „Listicle“) und

Tabelle 4.4: Verteilung der annotierten Klassen pro Annotator

	Annotator 1	Annotator 2	Annotator 3	Durchschnittlich
Nicht	1553	1158	1337	1349
Schwach	877	550	922	783
Medium	480	843	554	625
Stark	82	441	179	234

Tabelle 4.5: Absolute Übereinstimmungen bei balancierter Binarisierung

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	2992 (100%)	2043 (68%)	2373 (79%)
Annotator 2	2043 (68%)	2992 (100%)	2164 (72%)
Annotator 3	2373 (79%)	2164 (72%)	2992 (100%)

Annotationen zu vergleichen. Dieses ist in Abbildung 4.2 gezeigt.

4.4 Analyse der Annotationen

Es wurde bei der Annotation durch drei Annotatoren ein Fleiß-Kappa von 0.23 erreicht. Dies gilt als „Fair Agreement“, also eine „anständige“ Übereinstimmung. Krippendorffs Alpha beträgt ebenfalls 0.23. Es gibt eine durchschnittliche Übereinstimmung zwischen je zwei Annotatoren von 47.5%. In 861 Fällen (28.78%) waren sich alle Annotatoren einig. Die Anzahl der Klassen pro Annotator sind in Tabelle 4.4 zu finden. Annotator 2 hat dabei deutlich mehr Tweets als Mittlerer- oder Starker-Clickbait bewertet, als die anderen Annotatoren. Annotator 1 hat am wenigsten Tweets als Kein- oder Schwach-Clickbait gewertet.

Für unsere Arbeit haben wir für den Großteil unserer Experimente die Annotation binarisiert. Bei ausbalancierter Binarisierung, bei der je Nicht- und Schwacher-Clickbait zu „Nicht-Clickbait“ und Medium und Stark zu „Clickbait“ zusammengefasst werden, ergibt sich ein Fleiß-Kappa Wert von 0.35 und ein Krippendorffs Alpha von 0.35. Dabei stimmen alle Annotatoren in 2087 Fällen (69.75%) in ihrer Clickbait-Bewertung überein. Die durchschnittliche Übereinstimmung zwischen zwei Annotatoren beträgt 79.8%. Die Übereinstimmungsmatrizen für allgemeine Übereinstimmung und Clickbait zwischen den Annotatoren sind in den Tabellen 4.5 und 4.6 dargestellt.

In Abbildung 4.6 ist zu sehen, dass Annotator 1 115 Tweets als Clickbait bewertete, bei denen Annotator 3, der am meisten Clickbait fand, nicht in der Bewertung Clickbait übereinstimmt. Dies verdeutlicht, dass der große Unterschied in der Anzahl der als Clickbait bewerteten Tweets zwischen Annotator

Tabelle 4.6: Absolute Übereinstimmungen für Clickbait

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	562 (100%)	447 (34%)	338 (46%)
Annotator 2	447 (79%)	1281 (100%)	593 (80%)
Annotator 3	338 (60%)	593 (46%)	733 (100%)

1 und 2 nicht nur aus einer konservativeren, beziehungsweise liberaleren, Bewertung heraus entsteht, sondern dass es auch Gegensätze und strukturelle Unterschiede in der Kategorisierung gibt.

In einer auf den Annotationsprozess folgenden Diskussion ergab sich, dass Annotator 2 sensationalistische und wertende Ausdrücke strenger bewertete und daher eher eine Köderwirkung bei den Tweets feststellte als die anderen Annotatoren. Annotator 1 hingegen bewerte fehlende Themengenauigkeit und mangelnde Informationen strenger als die anderen Annotatoren. Annotator 3 war grundsätzlich konservativ in der Clickbait-Bewertung, hatte aber ausgewogene Schwerpunkte in seiner Bewertung.

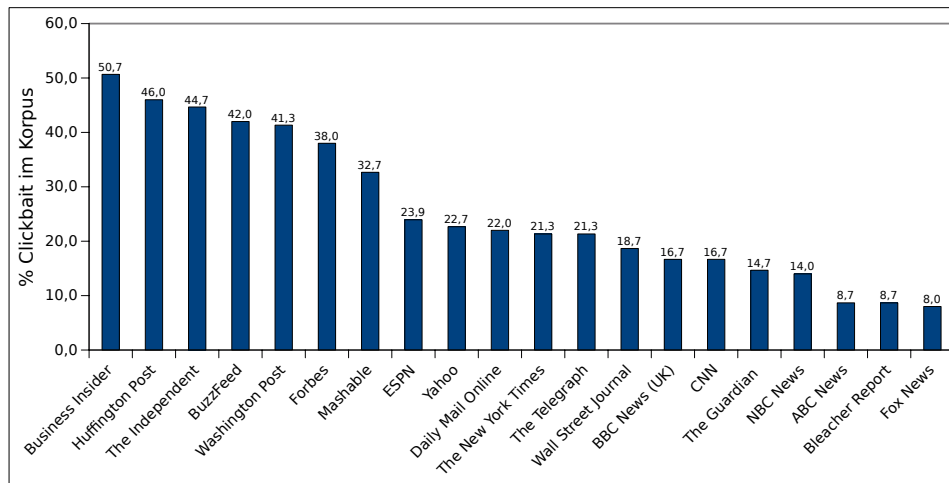
Letztendlich lässt sich keine abschließende Wahrheit über einen Tweet feststellen, da die Bewertung selbst bei vorheriger Abstimmung der Annotatoren nicht eindeutig ist. Diese Unterschiede, bei mit dem Thema vertrauten Personen, zeigen, dass Clickbait ein durch die Vielzahl der zusammenwirkenden Merkmale und die Komplexität von der in News-Überschriften verwendeten Sprache schwierig zu bewerten ist. Es lässt sich folgern, dass sich die empfundene Köderwirkung von Tweets von Person zu Person unterscheidet. Wir haben daher über einen Mehrheitsentscheid festgelegt, welche Tweets von uns als Clickbait betrachtet. Da die Übereinstimmung der Annotatoren bei binarisierter Annotation deutlich größer ist, haben wir uns dazu entschieden, diese im Weiteren zu verwenden.

Aus den Ergebnissen lässt sich schließen, dass es deutliche Unterschiede in der persönlichen Bewertung von Clickbait gibt. Wir stellen die These auf, dass Clickbait-Erkennung möglich ist, jedoch im Bestfall persönlich stattfinden sollte, um auf individuelle Präferenzen und Schwerpunkte eingehen zu können. Daher gehen wir auf diesen Faktor in unseren Experimenten ein.

Mehrfaches Erscheinen von Tweets

Während des Annotationsprozesses hat sich gezeigt, dass die Publisher des öfteren die gleichen Inhalte mehrfach getweetet haben. Wir gehen davon aus, dass die Herausgeber Tweets erneut veröffentlichen, um Leser mehrerer Zeit-zonen zu erreichen. Insgesamt gibt es 167 Tweets, deren Text mehr als einmal in unserem Korpus vorkommt.

Abbildung 4.3: Clickbait je Herausgeber



Im Folgenden wollen wir untersuchen, wie konsequent die Annotatoren mehrfach vorkommende Tweets annotiert haben. Dazu prüfen wir bei duplizierten Tweets die Konsistenz der Annotatoren, indem wir den Anteil der geänderten Meinungen bei binarisierter Betrachtung auswerten. 72 Tweets kamen dabei genau zwei mal vor, von diesen haben Annotator 1 und 2 bei 10 (13.9%) ihre Meinung von Clickbait zu Nicht-Clickbait, oder anders herum, geändert, Annotator 3 bei 14 (19.5%). Es ist somit erkennbar, dass die Bewertung von Clickbait deutlichen Schwankungen bei erneuter Bewertung eines bereits gesehenen Tweets unterliegt.

Erklärungsansätze für diese Diskrepanz ist zum einen, dass das vorher gesehene Tweets die Meinung über einen Tweet beeinflussen und verändern. Andererseits ist es Möglich, dass sich im Laufe des Annotationsprozesses die Vorstellung wie Clickbait gestaltet ist, ändert, da der Annotator eine größere Vergleichsmenge gesehen hat.

4.5 Clickbait-Verteilung pro Herausgeber

Es soll im Folgenden untersucht werden, ob sich die Publisher in der veröffentlichten Menge an Clickbait unterscheiden und falls ja, wie groß diese Unterschiede sind. Für die Betrachtungen wurden die Einzelentscheidungen der Annotatoren mithilfe der erwähnten balancierten Binarisierung auf Clickbait und Nicht-Clickbait vereinfacht. Daraufhin wurde die endgültige Klasse über einen Mehrheitsentscheid festgelegt. Bei dieser Betrachtung sind 25.64% der untersuchten Tweets Clickbait. Es lässt sich erkennen, dass die Publisher Business Insider (50.67%), Huffington Post (46%) und The Independent (44.67%)

Beispiel Tweets

Abbildung 4.4: Durch Bild Nicht-Clickbait



Abbildung 4.5: Durch Bild Clickbait



den meisten Clickbait auf Twitter veröffentlichen. ABC News (13%), Bleacher Report (13%) und Fox News (12%) veröffentlichten am Wenigsten. Dies zeigt große Unterschiede zwischen den Herausgebern, so hat Business Insider im Vergleich zu Fox News mehr als vier mal so viel Clickbait in unserem Korpus veröffentlicht. Wir untersuchen daher in Kapitel 6.2, inwiefern sich diese starken Differenzen auf die automatische Erkennung von Clickbait auswirken.

Einfluss von Bildern auf die Clickbait-Klasse

Einige Tweets des Korpus weisen für die Bestimmung der Clickbait-Stufe relevante Besonderheiten in der Verwendung von Medien auf. Wir werden diese im Folgenden näher betrachten, um die Komplexität des Annotationsvorganges zu verdeutlichen und um zu zeigen, welche Probleme bei der automatischen Erkennung zu erwarten sind.

Den in Abbildung 4.4 gezeigten Tweet würde dem Clickbait zugeordnet werden, vorausgesetzt, das dazugehörige Bild ist nicht zu sehen. Der Text ist vorwärtsreferenzierend und sehr wertend. Da „die vielleicht beste Tattoo-Vertuschung aller Zeiten“ direkt im Tweet zu sehen ist, kann dieser Tweet jedoch nicht als Clickbait gewertet werden. Entsprechend haben zwei Annotatoren dieses Bild als Nicht-Clickbait eingestuft und einer als schwaches Clickbait. Jedoch ist anzunehmen, dass dieser Tweet aufgrund seiner, für Clickbait

typischen, Wortwahl („this“) als Clickbait eingestuft wird. In dem Korpus sind einige derartige Tweets enthalten.

Als entgegengesetztes Beispiel sei der in Abbildung 4.5 gezeigte Tweet genannt. Während bei diesem Tweet die Überschrift sehr sachlich gehalten ist, wird durch das Verzichten der eigentlichen Selbstfotografie und das Verlinken eines nicht relevanten Bildes Clickbait erzeugt. Daher sind sich die Annotatoren bei dieser Tweets mit der Bewertung *Medium-Clickbait* einig.

Der Einfluss von Medien ist ein die Erkennung von Clickbait erschwerender Faktor, wie wir auch in Kapitel 6.4 zeigen.

4.6 Tweetlänge

Wir konnten feststellen, dass Clickbait-Tweets meist einen kürzeren Text haben. So haben die 2225 Clickbaits eine durchschnittliche Zeichenlänge von 70.5, wohingegen die als Nicht-Clickbait gewerteten eine durchschnittliche Länge von 81.1 Zeichen aufweisen. Wir gehen auf diesen Sachverhalt in unserem Modell ein. Zu beachten ist, dass Twitter als Korpusquelle besonders ist, da es nicht möglich ist, Nachrichten mit mehr als 140 Zeichen zu versenden. Weil jedoch auch Nicht-Clickbait nur in Ausnahmefällen diese Länge erreichen, ist davon auszugehen, dass dieses Kriterium auch bei anderen News-Aggregatoren als Twitter seine Gültigkeit behält.

4.7 Vorwärtsrefrenzierung

Blom und Hansen verwendeten in ihrer Arbeit zur Clickbait-Erkennung die dänischen Worte für „Hier“, „Diese“ und „Wir“. Wir haben das Vorkommen der englischen Entsprechungen „here“, „this“ und „we“ in den Tweet-Texten untersucht. Wir fanden Clickbait 317 Tweets bei denen diese Worte, unabhängig von Groß- oder Kleinschreibung, vorkamen, 156 (49%) wurden von der Mehrheit der Annotatoren als Clickbait gekennzeichnet, 161 (51%) als Nicht-Clickbait. In den 185 Tweets in denen „this“ vorkamen waren 60% Clickbait, bei „we“ nur 33.3% von 48, bei „here“ 31.2% von 96. Die von Blom und Hansen gewählten Worte sind als Clickbait-Identifikatoren somit nur eingeschränkt und bei getrennter Betrachtung geeignet.

4.8 Listicles

Wir haben eine einfache textbasierte Suche durchgeführt, bei welcher wir in den Texten der Tweets nach Zahlen und Zahlenwörtern (one, two, three, ...)

gesucht haben. Dabei haben wir in unserem Korpus 114 Listicles gefunden, von denen 90% als Clickbait eingestuft werden. Dies zeigt, dass dieses Format vor allem zur Verbreitung von Clickbait genutzt wird. Beliebt waren dabei Reise- und Essensthemen.

4.9 Andere Formen von Clickbait

Videoinhalte konnten in dieser Arbeit nicht tiefer ausgewertet werden, da uns keine Möglichkeiten diese in Attributen widerzuspiegeln, zur Verfügung standen. Es wurden 249 auf Videos verweisende Überschriften identifiziert, von welchen 57 (22.9%) in der der Mehrheitsentscheidung als Clickbait bewertet wurden. Mit einem Anteil von 8.4% an unserem Korpus stellen Video-Überschriften somit einen relevanten Teil der Tweets dar. Mit einem Clickbait-Anteil von etwa einem Viertel entspricht dieser dem des restlichen Korpus.

In unserem Korpus sind 13 Quiz-Überschriften enthalten, von diesen wurden in der Mehrheitsentscheidung der Annotatoren 10 als Clickbait gekennzeichnet. Aufgrund dem seltenen Vorkommen von Quiz-Clickbait in unserem Korpus konnten wir auf diese Form von Clickbait nicht detaillierter eingehen.

Kapitel 5

Clickbait-Erkennungsmodell

In diesem Kapitel werden wir das Modell, welches wir verwenden um Clickbait abzubilden, beschreiben. Mit Modell bezeichnen wir die Kombination aus Vorverarbeitung des Datensets, die verwendeten Attribute, Attributauswahl und die Klassifikatoren.

5.1 Attribute

Um unsere Attribute logisch zu unterteilen, haben wir einige Attributgruppen gebildet. Alle Attributgruppen werden in den Maschinelles-Lernen-Verfahren gleich behandelt. Die Gruppen werden in den folgenden Unterkapiteln jeweils tiefer behandelt. Die Hauptgruppen bilden dabei:

- Tweet
- Link
- Meta

Es wurden noch weitere Attributgruppen gebildet, diese dienen insbesondere dazu, Attribute zu kapseln, welche von Drittquellen abgeleitet wurden.

- Inquirer-Wortlisten: Entstanden aus den General-Inquirer-Daten
- Downworthy: Entstanden aus den Phrasen und regulären Ausdrücken des Browser-Plugins Downworthy

Es wurde keine Lemmatisierung, das Rückführen eines Wortes auf dessen Wortbasis, verwendet. Wir haben uns zu diesem Schritt entschieden, da sich in Experimenten innerhalb der Pilotkorpusphasen gezeigt hat, dass sich dies nicht positiv auf die Ergebnisse auswirkt. Des Weiteren konnten wir nicht ausschließen, dass Unterschiede in den Wortformen Informationen zur Erkennung von

Clickbait enthalten. Es wurden keine Stoppwörter entfernt. Dies sind Wörter, wie „a“, „actually“ etc., die sehr oft auftreten und oft keinen Informationsgehalt haben. Da wir nicht ausschließen konnten, dass deren Vorkommen für die Clickbait-Erkennung vorteilhaft ist, haben wir diese erhalten und ein Attribut implementiert, welches auf diese eingeht.

Bei der automatischen Berechnung einiger der verwendeten Werte lassen sich Fehler nicht vermeiden. So kann es sein, dass die Anzahl von Silben, Sätzen oder Wörtern, die für einige Attribute benötigt wird, trotz der Verwendung von Standardmethoden Abweichungen von manueller Berechnung aufweisen.

Tweet Attribute

Ziel dieser Attributgruppe ist es, den Tweet abzubilden. Dabei spielen einige Faktoren eine Rolle, etwa der Text selber und beigefügte Medien. Wir haben jedoch auch versucht, auf linguistische und psychologische Besonderheiten von Clickbait einzugehen. Einige Attribute, etwa mittlere Wortlänge, Flesch-Lesbarkeit und Wortlänge, wurden gewählt, weil wir die Vermutung hatten, dass Clickbait strukturell simpler ist als normale Tweets. Manche Attribute, etwa Anzahl Punkte und Zahlen wurden gewählt, um offensichtliche Besonderheiten von Clickbait darzustellen. So beginnen viele Listicles mit einer Zahl.

Bag-of-Words

Aus dem Text des Tweets wird ein Wordvektor gebildet, der anzeigt, welche Wörter, Wortketten, Zeichen und Zeichenketten dieser enthält. Dazu wurde die Bibliothek *aitools* verwendet, welche von der Fakultät *Webtechnologien und Informationssysteme* an der Bauhaus-Universität Weimar entwickelt wurde. Der Text des Tweets wird mittels des eines Tokenisierers in Wort- und Zeichen-1-, 2- und 3 -Gramme zerlegt. Für jedes entstandene n-Gramm wird gezählt, wie oft es in dem Text des Tweets vorkommt.

Abbildung 5.1: Beispiel Bag-of-Words

Zeichen			Wörter		
1	2	3	1	2	3
F	Fa	ace	Families	Families who	Families who have
a	al	ace	a	a parent	a parent coming
c	am	ami	coming	coming out	coming out share
d	ar	are	faced	faced the	faced the fallout
e	av	ave	fallout	fallout of	fallout of a
...					

Part-of-Speech Tagging (POST)

POST bezeichnet das Zuordnen von Wörtern zu ihrer grammatikalischen Funktion. Wir haben hierfür das Java-Framework Stanford CoreNLP verwendet[14]. Da dessen Standardmodell für Twitter-Nachrichten nicht optimiert ist, haben wir ein von Derczynski et al. erstelltes, speziell für Twitter geeignetes Modell verwendet[43]. Dies soll an folgendem Beispiel verdeutlicht werden, bei welchem folgende Abkürzungen verwendet werden: VB= Verb, Basisform; TO= zu; PRP\$= Personalpronomen; NN= Nomen; VBD= Verb, Vergangenheit. Im Anhang unter A.1 werden weitere vom Stanford CoreNLP Framework verwendete Part-of-Speech Tags aufgelistet.

Abbildung 5.2: Beispiel Annotation POS Tagger

Prepare	to	have	your	inbox	carpet	bombed
VB	TO	VB	PRP\$	NN	NN	VBD

Es werden aus den POS Tags Mono-, Bi- und Trigramme gebildet. Deren Vorkommen wird, genauso wie im Bag-of-Words, gezählt, so dass in diesem Beispiel, unter anderem: „NN 2“, „PRP\$ NN 1“ und „VB PRP\$ NN 1“ gebildet werden.

Bindungskraft, Erregung und Dominanz

Um die in Kapitel 2.4 erwähnten VAD-Skalen abzubilden, wurde je ein Attribut für Bindungskraft, Erregung und Dominanz gebildet. Dazu wurden die jeweiligen Werte (in dem Datenset die „Mean.Sum“ Spalte) verwendet. Um die VAD-Werte für einen Tweet zu bilden, wurde nach in dem Datenset vorkommenden Wörtern gesucht und über diese der Mittelwert gebildet. Da diese Skalen über den Wertebereich 1-9 definiert sind, wurde bei Tweets ohne Wörtern mit bekannten VAD-Werten der Mittelwert 5 der Skala angenommen.

Imagga Bildtags

Diese Attribute dienen dazu, den Inhalt von Bildern in Tweets zu beschreiben. Dabei wurden 3477 einzigartige Bildtags erzeugt, wobei einen Tweet, der ein Bild enthält, durchschnittlich 58.77 Bildtags hat.

Abbildung 5.3: Imagga Bildtags



- building
- architecture
- house
- city
- home
- ...

Länge des längsten Wortes

Gibt die Zeichenlänge des längsten Wortes an.

„Families who have faced the fallout of a
parent coming out share their stories“ 7

Mittlere Wortlänge

Gibt die mittlere Wortlänge an.

„RT @BuzzFeedCeleb: The ladies of OITNB
slayed at OrangeCon“ 5.5

Stoppwörter

Dieses Attribut an wie groß der Anteil der „Inhaltsleeren“ Wörter an dem Tweettext ist. Stoppwörter sind Wörter, die oft auftreten, aber selbst keine Wichtigkeit haben[44]. Es gibt keine einheitliche Liste dieser Wörter, für unsere Untersuchung wurde eine Liste¹ gewählt die 671 Wörter enthält, darunter Beispielsweise „to“, „hers“ und „furthermore“. Um das Auftreten von Stoppwörtern relativ zu beschreiben, haben wir das Verhältnis von der Anzahl von Stoppwörtern zur Anzahl aller Wörter gebildet. Die Stoppwörter sind in den folgenden Beispielen hervorgehoben.

„ This *little* boy accidentally told *his* mom *what*
every parent *needs* to hear“ 0.46

¹Die Wörter wurden folgender Quelle entnommen: code.google.com/archive/p/stop-words/downloads Stand: 15.06.2015

Einfache Wörter

Dazu wurde eine von Dale und Chall erarbeitete Liste² einfacher Wörter verwendet. Diese besteht aus 3000 Wörtern, die von amerikanischen Viertklässlern meistens verstanden werden[45]. Wir verwenden das Verhältnis zwischen der Anzahl der Stoppwörter zur Gesamtanzahl der Wörter.

Flesch-Lesbarkeit

Die Flesch-Lesbarkeit ist ein von Flesch entwickelte Wert, die die Lesbarkeit eines Textes berechnet[46]. Sie ermöglicht es, die Lesbarkeit eines Textes auf eine Zahl zu reduzieren.

$$206.834 - 1.015 \frac{\text{Anzahl Wörter}}{\text{Anzahl Sätze}} - 84 - 6 \frac{\text{Anzahl Silben}}{\text{Anzahl Wörter}}$$

„This little boy accidentally told his mom what every parent needs to hear“	76.5
--	------

Anzahl Punkte

Die Anzahl der Punkte in dem Text des Tweets. Bei diesem Attribut wurden auch Sonderfälle, etwa das drei Punkte Unicode Zeichen³, welches verwendet wird um die Zeichenanzahl in einem Tweet gering zu halten, einbezogen.

„RT @DailyMailCeleb: Foo Fighter’s rocker Dave Grohl breaks his leg on stage...but just keeps on singing...“	6
--	---

Stimmungsanalyse

Hier wird ebenfalls das Stanford CoreNLP verwendet. Dieses berechnet die Stimmung jedes Satzes[14]. Dabei wird zwischen *very negative*, *negative*, *neutral*, *positive* und *very positive* unterschieden. Wir verwenden die Stimmung des längsten Satzes.

„Bullied teen creates magnificent prom dress and it is perfect payback“	very positive
--	---------------

²Gefunden unter: countwordsworth.com/download/DaleChallEasyWordList.txt
Stand: 29.03.2016

³U+2026

Hashtags

Hier werden jeweils die in dem Tweet vorkommenden Hashtags verwendet. Ein weiteres Attribut gibt an, ob der Tweet ein Hashtag enthält.

„#FIFA suspends bidding for 2026 World Cup amid corruption scandal“	#FIFA
--	-------

Erwähnte Nutzer

Hier werden (Twitter-)Nutzer, die in einem Tweet erwähnt werden, angegeben.

„RT @DailyMailCeleb: Foo Fighter’s rocker Dave Grohl breaks his leg on stage...but just keeps on singing...“	@DailyMailCeleb
--	-----------------

Anzahl Zahlenzeichen

Zählt die Anzahl der auftretenden Zahlenzeichen. Dabei wird keine Rücksicht auf deren Kontext genommen.

„66 thoughts I had while watching season 5, episode 9 of Game Of Thrones“	4
--	---

Beginnt mit Zahl

Gibt an, ob ein Tweet mit einem Text beginnt. Dabei wird in dem Fall, dass der Tweet selbst retweetet wurde, der Originaltext verwendet.

„RT @komonews: 2 anti-drilling activists hanging from Shell ship in Bellingham“	yes
--	-----

Weitere

Um verschiedene Aspekte abstrakt abbilden zu können, haben wir mehrere Wortlisten verwendet. Es wurde aus jeder Wortliste ein Attribut gebildet, welches angibt, wie viele Wörter der Liste in dem jeweiligen Tweet auftreten.

- Pronomen: Hier wurden 73 englische Pronomen verwendet. Die Liste wurde aus enchantedlearning.com/wordlist/pronouns.shtml⁴ extrahiert.

⁴Stand: 29.03.2016

- Präpositionen: 70 englische Präpositionen, entnommen aus: englishclub.com/grammar/prepositions-list.htm Stand: 29.03.2016
- Klischees: englische Klischees oder Redewendungen, Beispielsweise „the early bird gets the worm“ oder auch kürzer „God help us“. Diese Liste hat 1659 Einträge, wird jedoch de facto nicht verwendet, da weniger als 1% der Tweets in dieser Liste enthaltene Klischees verwenden.
- *To be Verben*: Enthält 11 Konjugationen des englischen Verbs *sein*.

Verlinkter Inhalt

Bei der Betrachtung des verlinkten Inhalts haben wir ausschließlich dessen Hauptinhalt betrachtet. Hauptinhalt einer Internetseite ist dabei der Text, der den tatsächlichen Inhalt der Seite darstellt, ohne Menüs, HTML- oder JavaScript-Elemente und Ähnlichem. Um den Hauptinhalt des verlinkten Inhalts zu erhalten, haben wir die Java-Bibliothek *Boilerpipe* verwendet, welche diesen anhand verschiedener Parameter extrahiert. Diese Attributgruppe enthält nicht alle für den Tweet eingesetzten Attribute, da wir den Tweet als deutlich maßgebender für die Clickbait-Erkennung erachten.

Bag-of-Words

Dieser wird aus dem Haupttext des verlinkten Inhaltes gebildet und funktioniert genau so wie der Bag-of-words aus Kapitel 5.1.

Flesch-Lesbarkeit

Wie bei Tweet-Attributgruppe.

Zeichen- und Wortlänge des Hauptinhalts

Wie bei Tweet-Attributgruppe.

Meta Attribute

Diese Attributgruppe enthält Attribute die den Tweet nicht direkt beschreiben, sondern den Kontext eines Tweets widerspiegeln.

Ersteller

Gibt den Ersteller des Tweets, in unserem Fall immer ein News-Herausgeber, an.

Medientyp

Gibt den Typ des anhängenden Mediums an (Bild, Video), falls der Tweet Medien enthält. Falls keine Medien angehängt sind, wird dies gekennzeichnet.

Tageszeit

Die Zeit, in welcher der Tweet abgeschickt wurde. Dies wird in vier Tageszeiten aufgeteilt: Morgen, Mittag, Abend und Nacht. Dafür wird die von der Twitter-API zur Verfügung gestellte Zeit verwendet, die Weltzeit UTC. Da der Tweet in unserem Korpus jedoch meist keinen Ort der Erstellung enthält, konnten wir die lokale Uhrzeit der Tweeterstellung nicht direkt ermitteln.

Teilungen

Dieses Attribut gibt die Anzahl der „shares“ auf Twitter an. Dazu wurde die Anzahl, die zum Zeitpunkt der Korpuserstellung vorhanden war, genutzt.

Retweet

Gibt an, ob der Tweet selbst retweetet wurde. Dies machen die Herausgeber, um Artikel aus ihren Unterabteilungen (z.B: Sport, Kochen, Entertainment) auf ihrem Haupt-Twitteraccount hervorzuheben. Diese Information kann entweder durch Test, ob ein Tweet mit „RT“ beginnt, oder über ein Attribut des JSON-Objektes erlangt werden. Wir haben uns für letztere Möglichkeit entschieden.

Textwiederholungen

Dieses Attribut gibt an, wie oft ein Text im Korpus vorkommt (siehe Abschnitt 4.4).

Downworthy

Für die Attributgruppe wurde das Downworthy-Browser-Plugin ausgewertet. Dieses verwendet eine Reihe von Regexen und Ausdrücken („WOW“) um Clickbait zu erkennen, um daraufhin diese Redewendungen durch Ersetzen mit anderen Ausdrücken lächerlich zu machen. So wird aus „Will Blow Your Mind“, einem typischen Clickbait-Ausdruck, „Might Perhaps Mildly Entertain You For a Moment“. Es wurden sechs Regexe und 62 Ausdrücke aus dem Quellcode des Plugins extrahiert. Da das Browser-Plugin in JavaScript implementiert ist, wurden die Regexe für Java übertragen. Ein Attribut gibt an, ob der Text des

Abbildung 5.4: Beispiel Downworthy Regex

RT @fxnopinion: Ian Bremmer: Ameri-
(.*) (Is | Can | Do | Will) (.*) \\ ? ca badly needs a coherent foreign poli-
cy strategy. Will 2016 candidates ever
share their vision? ...

Tweets einem der gefundenen Regexe entspricht und ein Attribut, welches angibt, wie viele der Ausdrücke vorkommen. Ein Beispiel für verwendete Regexe ist in Abbildung 5.4 gegeben. Für jeden Regex wird überprüft, ob der Text des Tweets auf diesen zutrifft. Im Beispiel ist der auf den Regex zutreffende Text hervorgehoben.

Inquirer-Wortlisten

Eine weitere Attributkategorie stellen die aus den *General Inquirer* Daten erzeugten Wortlisten-Attribute dar. General Inquirer ist ein computergestützter Ansatz zur Inhaltsanalyse von Text-Daten[47]. Die Listen wurden durch Auswertung der verfügbaren online Ressourcen erzeugt.⁵ Darin sind Beispielsweise der Wortgruppe „You“ die Wörter „thee“, „thou“, „thy“, „you“, „you“, „your“, „yours“, „yourself“ und „yourselves“ zugeordnet. Es entstanden 182 Wortlisten, wobei aus jeder Wortliste ein Attribut entstanden ist welches beschreibt, wie viele Worte aus der jeweiligen Liste in dem Tweet-Text gegeben sind.

5.2 Vorverarbeitung

In einem ersten Aufbereitungsschritt wurde das Rohkorpus in ein Zwischenformat umgewandelt. Dabei wird der Tweet, sowie die assoziierten Dateien mittels der Features in Dateien umgewandelt, welche die extrahierten Attribute darstellen. Die entstehende Repräsentation enthält nur noch die Attribute des Tweets. Dies ermöglicht eine vereinfachte und schnellere Neubildung des Datensatzes mit anderen Parametern. Die entstehenden Attributdateien werden daraufhin zu einem Datenset in das Weka eigene Arff-Format umgewandelt und erhalten mittels einer gegebenen Annotation ihre Clickbaitklasse. In diesem Schritt werden die Tweets chronologisch sortiert. Das entstandene Datenset wird daraufhin in ein Trainings- und ein Testset aufgespalten. Wir haben bei den meisten Experimenten eine Aufteilung von $\frac{2}{3}$ zu $\frac{1}{3}$ verwendet. Das heißt, dass die chronologisch ersten zwei Drittel der Tweets das Trainingsset

⁵wjh.harvard.edu/~inquirer/spreadsheet_guide.htm Stand 10.11.2015,
wjh.harvard.edu/~inquirer/inqdict.txt Stand: 10.11.2015

bilden und das restliche Drittel als Testset dient. Um die Größe des Datensets zu beschränken und um *overfitting* zu vermeiden, werden Attribute, die in weniger als 1% der Datenpunkte mit einem positiven Wert auftreten, in den beiden Datensätzen entfernt. Overfitting bezeichnet zu spezifisches Training durch das Datenset, so dass zu viele Attribute in das Modell einbezogen werden, welches die Generalisierbarkeit des Modells schwächt und die Leistung der (Clickbait-)Erkennung reduziert[48].

Da der von uns erstellte Korpus kein ausgeglichenes Klassenverhältnis hat (nur etwa ein Viertel der Tweets sind Clickbait), haben wir uns dazu entschieden, *Oversampling* zu verwenden. Dieses bezeichnet ein Verfahren, bei welchem der Anteil der Klassen, die selten vorkommen, im Trainingsset künstlich erhöht wird, indem Instanzen der seltenen Klassen zufällig verdoppelt werden[49]. Somit konnte sicher gestellt werden, dass das Klassenverhältnis ausgeglichen ist.

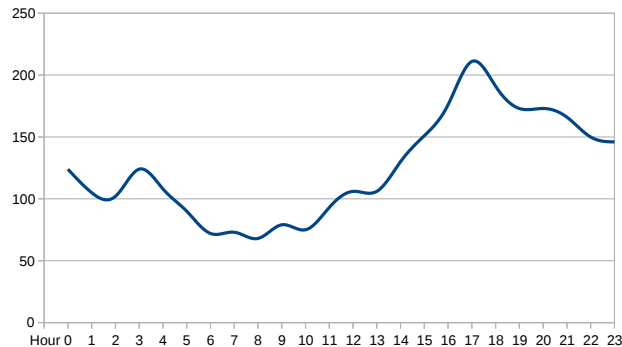
Diskussion

Einige der erwähnten Attribute sind durch die in Abschnitt 5.2 beschriebene Vorauswahl durch Mindestvorkommen nicht im tatsächlich verwendeten Datenset aufgetreten, wurden jedoch der Vollständigkeit halber erwähnt. So wurden zu wenige Downworthy-Ausdrücke gefunden, die Attributgruppe besteht lediglich aus dem Downworthy-Regexen und auch diese wurden nur in sehr wenigen Tweets gefunden. Das Abbildung 5.4 gegebene Beispiel ist der einzige Regex, zu dem in dem Korpus zutreffende Texte erkannt wurden. 53 (7%) Tweets im Korpus haben einen auf den Regex passenden Text, davon sind alle in der Mehrheitsentscheidung Clickbait. Dieses Attribut ist, falls ein Ausdruck gefunden wird also ein sehr guter Indikator für Clickbait. Da das Plugin weder für eine wissenschaftliche Arbeit entwickelt noch getestet wurde, sondern eher ein Witz, beziehungsweise ein *Proof-of-concept* konzipiert wurde, ist davon auszugehen, dass die Leistung dieser Attributgruppe, aufgrund des vergleichsweise seltenen Auftretens der Wendungen, schlecht ist. Da die Downworthy-Attribute jedoch den einfachen Ansatz zur Clickbait-Erkennung darstellt, bei welchem eine Reihe typischer Wendungen direkt, oder als Regexe dargestellt, verwendet werden, stellt diese Attributgruppe die Basislinie zur Bewertung anderer Attribute dar.

Erwähnte Nutzer ist ebenfalls im Datenset praktisch nicht enthalten, da nur ein Nutzer in mehr als 1% der Tweets erwähnt wird und somit nach dem in Kapitel 5.2 beschriebenen preprocessing im Datenset verbleibt. Ebenfalls nicht im letztlich resultierenden Datenset waren unter anderem Tweet-Wort-Trigramme.

Insbesondere das Meta-Attribut *Ersteller* ist im Kontext der Clickbait-Erkennung differenziert zu betrachten. Hier stellt sich die Frage, ob dieses verwendet werden sollte, da dies zur Folge hat, dass weniger der Tweet selbst

Abbildung 5.5: Anzahl der pro Stunde veröffentlichten Tweets in UTC



bewertet wird, sondern vielmehr ein Vorurteil aufgrund der vorherigen Tweets des Erstellers gebildet wird. Zum einen ist es nachvollziehbar, diese zu verwenden, da in dem hypothetischen Fall, dass ein Ersteller ausschließlich Clickbait verbreitet, davon auszugehen ist, dass auch weitere Tweets von diesem Ersteller ebenfalls Clickbait sind. Auch lässt sich dieses Attribut als „Sicherheitsnetz“ verstehen, welches die Einschätzung des Klassifikators korrigiert, falls der Tweet selbst durch unübliche Stilmittel schwer zu bewerten ist. In diesem Fall würde dieses Attribut dazu dienen, die Bewertung auf den, für den Ersteller, wahrscheinlichsten Fall anzupassen. Andererseits ist es für die akademische Zielsetzung *automatische Clickbait-Erkennung* passender und anstrebenswerter, sich auf stilistische Attribute zu konzentrieren, die möglichst unabhängig von Ersteller und Tagesthemen sind, so dass das gebildete Modell universell und anpassungsfähig wird. Um sich der dieser Diskussion zu Grunde liegende Frage zu widmen, wie groß der Einfluss der Ersteller eines Tweets auf die Leistung des Modells ist, haben wir in Kapitel 6.2 weitere Experimente durchgeführt. Dabei hat sich ergeben, dass das Attribut *Ersteller* zwar einen positive Wirkung auf die Clickbait-Erkennung hat, der Einfluss jedoch gering und kompensierbar ist.

Wie in Abbildung 5.5 zu sehen, lässt sich aber ein deutlicher Anstieg der Tweets zwischen 12 bis 17 Uhr feststellen, was darauf schließen lässt, dass es Stoßzeiten für Nachrichten-Tweets gibt. Der Anteil an veröffentlichtem Clickbait von allen Nachrichten bleibt über den Tag konstant.

5.3 Attributauswahl

Attributauswahl bezeichnet verschiedene Verfahren, bei welchen die Anzahl der Attribute im Datenset reduziert wird. Dies hat viele Vorteile, so reduziert es die zum Lernen benötigte Zeit des Klassifikators und es werden Attribute

entfernt, die nicht mit den Klassenattributen in Zusammenhang stehen oder irreführend sind. Die Attributauswahl verbessert die Leistung der Klassifikatoren teils beträchtlich. Um die Attributauswahl durchzuführen, verwenden wir das χ^2 -Verfahren.

Bei dem χ^2 -Verfahren wird die statistische Abhängigkeit zwischen Attribut und den Klassen, in unserem Fall Clickbait oder Nicht-Clickbait, berechnet. Es wird für jedes Attribut der χ^2 -Wert errechnet, wobei ein hoher Wert angibt, dass die Null-Hypothese, die behauptet dass zwei Ereignisse unabhängig sind, verworfen werden muss. Dies heißt, dass bei einem hohen χ^2 -Wert eine große statistische Abhängigkeit gegeben ist. Dabei werden für die reduzierte Attributmenge die Attribute mit der höchsten statistischen Abhängigkeit zur gesuchten Klasse ausgewählt.

5.4 Klassifikatoren

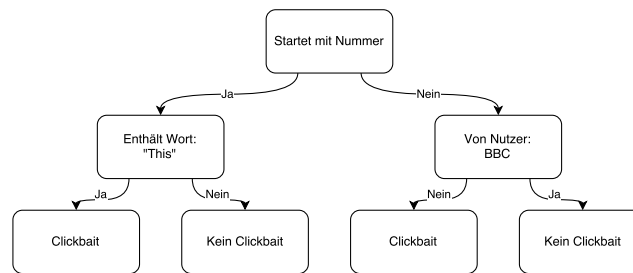
Klassifikatoren bezeichnen Algorithmen, welche einen Attributvektor klassifiziert und einer Klasse, hier der Clickbait-Klasse die Clickbait und Nicht-Clickbait enthält, zuordnet. Der Klassifikator wird zunächst mit einem Datenset trainiert, dieses wird meist als Trainingsset bezeichnet. Die Leistung seiner Vorhersagen wird gegen ein Testset verglichen. In unserer Untersuchung haben wir die vier Klassifikatoren Naiver Bayes, Random Forest, Logistische Regression und J48 verglichen, um den für die Clickbait-Erkennung am besten geeignetsten zu bestimmen. Im Folgenden werden wir die verwendeten Klassifikatoren beschreiben.

Naiver Bayes

Der Naiver-Bayes-Klassifikator (NB) ist ein Klassifikator, der auf dem Satz von Bayes basiert.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Bei dem Naiven Ansatz wird angenommen, dass jedes Attribut direkt von dem Klassenattribut abhängt und statistisch unabhängig von anderen Attributen ist. Er bestimmt die Klasse eines Datenpunktes, indem er in dem Trainingsdatensatz die bedingten Wahrscheinlichkeiten zwischen der zu bestimmenden Klasse und dem jeweiligen Attribut bestimmt[50]. Er zeichnet sich besonders durch seine Schnelligkeit aus.

Abbildung 5.6: Beispiel Entscheidungsbaum

J48

Der J48 ist eine Implementation des C4.5 Algorithmus von Weka und klassifiziert die Attributvektoren mithilfe eines Entscheidungsbaums[51]. Bei einem Entscheidungsbaum wird im Trainingsschritt eine Abfolge von Entscheidungen gelernt, die Knoten eines Baumes darstellen. Bei Bildung des Baumes wird jeweils diejenige Entscheidung gesucht und als neuer Knoten gewählt, die die größte Information enthält. Es werden weiterhin Verfahren angewendet, welche die Komplexität des Baumes begrenzen. In Abbildung 5.6 wird dargestellt, wie ein Entscheidungsbaum aussehen könnte.

Random Forest

Bei dem Random-Forest-Klassifikator (RF) werden mehrere kleine Entscheidungsbäume durch Auswahl von zufälligen Attributen gebildet[52]. Der Random Forest verwendet „Bagging“ für die einzelnen Bäume, bei welchem das Trainingsset in verschiedene Untersets, bei welchen die Klassenverteilung erhalten bleibt, für jeden Baum aufgeteilt wird. Der Klassifikator verwendet in unserer Konfiguration 100 Entscheidungsbäume mit je 7 Attributen. Das Verfahren bestimmt eine Klasse, indem jeder der entstandenen Bäume eine Entscheidung trifft und somit eine *Stimme* abgibt, der Klassifikator wählt daraufhin die Klasse mit den meisten Stimmen.

Logistische Regression

Der Logistische-Regression-Klassifikator (LR) verwendet zur Bestimmung des Klassenattributes eine logistische Regressionsformel[53]. Das heißt, dass die Klassen durch eine, von der Anzahl der Attribute abhängige, vieldimensionale logarithmische Funktion getrennt werden.

Abbildung 5.7: Terminologie Übersicht

		Erkannt		Recall
		Clickbait	Nicht-Clickbait	
Tatsächlich	Clickbait	true positive	false negative	
	Nicht-Clickbait	false positive	true negative	
		Precision		

5.5 Werte zur Leistungsmessung

Um die Leistung eines Modells zu analysieren, stehen einige Werte zur Verfügung[54]. In dieser Arbeit wollen wir uns jedoch auf drei grundlegende Werte beschränken, die es uns ermöglichen, die Leistung des Klassifikators differenziert zu analysieren. Diese sind:

- Recall
- Precision
- ROC-Area

Für die folgenden Erklärungen sind einige Begriffe nötig, welche hier kurz am Beispiel der Clickbait-Erkennung erläutert werden sollen und in Abbildung 5.7 dargestellt sind. Falls wir von Leistung des Klassifikators sprechen, ist dies allgemein, alle Faktoren einbeziehend, zu verstehen.

- *true positive* (TP): diejenigen Elemente, die vom Klassifikator korrekt als der gesuchten Klasse (in diesem Fall Clickbait/ Nicht-Clickbait) angehörig erkannt wurden. Die *true positive rate* (TPR) entspricht dabei $\frac{TP}{TP+FN}$.
- *true negative* (TN): diejenigen Elemente, die vom Klassifikator korrekt als nicht der gesuchten Klasse angehörig erkannt wurden.
- *false positive* (FP): diejenigen Elemente, die vom Klassifikator inkorrekt als der gesuchten Klasse angehörig erkannt wurden. Die *false positive rate* (FPR) entspricht dabei $\frac{FP}{FP+TN}$.
- *false negative* (FN): diejenigen Elemente, die vom Klassifikator inkorrekt als nicht der gesuchten Klasse angehörig erkannt wurden.

Recall

Recall (auch „Trefferquote“) berechnet sich durch[54]:

$$Recall = \frac{TP}{TP + FN}$$

Recall gibt an, wie groß der Anteil des gefunden Clickbait zu allen Clickbait-Tweets ist.

Precision

Precision (auch „Genauigkeit“) berechnet sich durch[54]:

$$Precision = \frac{TP}{TP + FP}$$

Precision gibt an, welcher Anteil der als Clickbait erkannten Tweets tatsächlich Clickbait ist.

Gewichteter Durchschnitt

Precision und Recall werden bei Angabe jeweils im Bezug auf die Klasse Clickbait und Nicht-Clickbait berechnet, es entstehen also immer zwei Werte für Recall und Precision. Wir verwenden, um diese Werte anzugeben, den gewichteten Durchschnitt, bei welchem die Anzahl der Klassen einer Instanz eingerechnet wird.

ROC-Area

Ein ROC-Graph ist ein Graph, bei welchem der Nutzen der Klassifizierung, in Form der TPR, den Kosten in Form der FPR gegenübergestellt wird[54]. Dabei steht ROC für „Receiver Operating Characteristics“. Es wird eine Kurve gebildet, welche angibt, welche FPR ein Modell bei einer gegebenen TPR hat. Eine perfekte ROC-Kurve verläuft über den Punkt (0,1), bei welchem der Klassifikator immer richtig und nie falsch bewertet. Die Fläche unter der Kurve wird als *Area under the curve* (ROC-Area) bezeichnet. Ein perfekter Klassifikator hat eine ROC-Area von 1.0. Das stetige auswählen der häufigsten Klasse des Trainingssets (bekannt als ZeroR Klassifikator) ergibt eine ROC-Area von 0.5, und stellt den praktisch schlechtesten Wert dar, da ein Klassifikator der eine ROC-Area von unter 0.5 erreicht, umgedreht werden kann, also die vorhergesehen Klassen getauscht werden um einen Wert von über 0.5 zu erreichen. Die ROC-Area ist ein zur Messung von Klassifikationsleistung besonders geeigneter

Wert, da dieser auch bei stark unbalancierten Klassenverteilungen gut funktioniert und da in ihm alle Schwellwerte des Klassifikators abgebildet werden[54]. Des weiteren ist er nicht Klassenbezogen, wie es Precision und Recall sind.

Kapitel 6

Evaluierung

Im Folgenden werden die verschiedenen von uns durchgeführten Experimentserien beschrieben und ausgewertet. Dazu wird zunächst das Hauptexperiment, welches die Erkennung von Clickbait untersucht, unserer Arbeit erläutert und analysiert. Daraufhin werden einige weiterführende Aspekte und Fragestellungen, die durch die durch Korpusanalyse und die Auswertung des Hauptexperimentes entstanden sind, in weiteren Experimenten untersucht.

6.1 Hauptexperiment

Zunächst werden wir die erste Fragestellung dieser Arbeit zu beantworten:

- Wie gut lässt sich ein allgemeines Clickbait-Modell erstellen?

Diese Experimentserie stellt den Kern dieser Arbeit da. Dazu wurden mehrere Schritte durchgeführt, die gewährleisten, dass diese Frage möglichst präzise beantwortet werden kann.

Es wurde dazu die Annotation jedes Bewerbers balanciert binarisiert. Das heißt, es wurden Tweets, die als Nicht-Clickbait oder Schwacher-Clickbait eingestuft wurden, als Nicht-Clickbait gewertet. Tweets, die als Medium-Clickbait oder Starker-Clickbait eingestuft wurden, wurden zu Clickbait. Es wurde daraufhin mittels Mehrheitsentscheid zwischen den Annotatoren die endgültige Clickbait-Klasse für jeden Tweet ermittelt. Somit wurde ein Tweet als Clickbait gewertet, falls mindestens zwei Annotatoren einen Tweet als Medium- oder Starken-Clickbait gewertet haben.

Die Tweets des Korpus wurden in chronologischer Ordnung verwendet. Dies ermöglicht es uns zu erkennen, wie gut der Klassifikator Tweets deren Inhalte, in Form von Wörtern, Hashtags, Personen und Ähnlichem, zuordnen kann ohne diese zu kennen.

Tabelle 6.1: ROC-Area Übersicht

	NB	RF	LG	J48
Alle Tweet Attr.	0.748	0.814	0.558	0.576
Alle Meta Attr.	0.639	0.61	0.674	0.606
Alle verl. Inh. Attr.	0.465	0.541	0.56	0.558
Downworthy	0.5	0.5	0.5	0.5
Inquirer	0.659	0.644	0.649	0.572
X^2 10 Attr.	0.746	0.731	0.775	0.65
X^2 100 Attr.	0.746	0.808	0.749	0.637
X^2 1000 Attr.	0.625	0.785	0.533	0.572
Alle Attr.	0.581	0.74	0.749	0.579

Das Korpus wurde daraufhin jeweils in ein Trainings- und ein Testset aufgeteilt, bei welchem das Trainingsset zwei Drittel der Tweets enthält und das Testset das verbleibende Drittel. Durch diesen Experimentaufbau sind keinerlei Informationen aus dem Testset im Trainingsset, da im Preprocessingschritt ausschließlich Attribute ausgewählt wurden, die im Trainingsset in mehr als 1% der Tweets vorkommen.

Da das Verhältnis von Clickbait zu Nicht-Clickbait etwa 1:4 beträgt, haben wir, um die Klassenverteilung auszugleichen, das in Kapitel 5.2 beschriebene *Oversampling* verwendet. Das Trainingsset enthält dadurch 987 Clickbait und Nicht-Clickbait Tweets.

Um Einfluss und Leistung einzelner Attribute und Attributgruppen nachvollziehen zu können, wurden die Klassifikatoren mit diesen trainiert. Dabei wurden von uns implementierte Attribute, etwa *Länge des Tweets in Zeichen* oder *Flesch-Lesbarkeit des Tweets* gesondert betrachtet. Größere Attributgruppen wie Inquirer-Wortlisten oder Zeichen n-Gramme hingegen wurden zusammengefasst. Bei der Auswertung dieses Experimentes ist zu beachten, dass diese Ergebnisse nur einen einmaligen Durchlauf darstellen und die Ergebnisse durch verschiedene Zufallsfaktoren, etwa durch Oversampling oder bei dem Trainieren des Klassifikators, abweichen können.

Um die Leistung der Klassifikatoren bei Attributauswahl zu überprüfen haben wir 10, 100 und 1000 Attribute aus der Menge aller Attribute durch das X^2 -Verfahren ausgewählt.

Dabei ergaben sich die in den Tabellen 6.1, 6.2, 6.3 und 6.4 dargestellten Werte.

Tabelle 6.2: ROC-Area Tweet-Attribute

	NB	RF	LR	J48
Alle	0.748	0.814	0.558	0.576
Anzahl Punkte	0.538	0.535	0.539	0.536
Einfache Wörter	0.645	0.588	0.645	0.619
Enthält Hashtag	0.528	0.528	0.528	0.528
Hashtags	0.504	0.504	0.504	0.504
Präpositionen	0.612	0.612	0.612	0.593
Pronomen	0.627	0.627	0.627	0.612
To Be	0.533	0.539	0.533	0.533
Bild-Tags	0.554	0.524	0.514	0.531
Max. Wortlänge	0.56	0.519	0.564	0.536
Mittlere Wortlänge	0.618	0.545	0.619	0.555
Erwähnte Nutzer	0.472	0.472	0.472	0.5
Zeichenlänge	0.649	0.573	0.649	0.592
Lesbarkeit	0.661	0.591	0.66	0.61
Beginnt mit Nummer	0.554	0.554	0.554	0.554
Wort- und Zeichen n-Gramme	0.732	0.786	0.53	0.594
Zeichen 1-Gramme	0.667	0.727	0.732	0.563
Zeichen 2-Gramme	0.727	0.761	0.574	0.586
Zeichen 3-Gramme	0.702	0.756	0.542	0.581
Zeichen n-Gramme	0.726	0.792	0.527	0.598
Wort 1-Gramme	0.716	0.689	0.723	0.648
Wort 2-Gramme	0.519	0.535	0.534	0.526
Wort 3-Gramme	0.5	0.5	0.5	0.5
Wort n-Gramme	0.719	0.684	0.72	0.651
POS-Tags	0.779	0.803	0.647	0.66
Stoppwörter	0.652	0.561	0.652	0.581
Sentiment	0.577	0.577	0.578	0.578
VAD-Erregung	0.523	0.501	0.552	0.498
VAD-Dominanz	0.563	0.543	0.563	0.544
VAD-Bindung	0.558	0.564	0.558	0.551

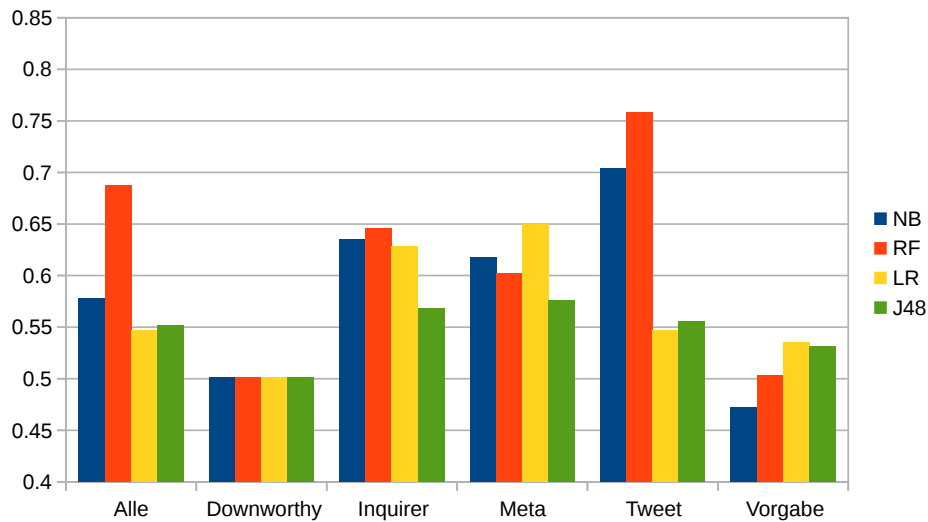
Tabelle 6.3: ROC-Area Meta-Attribute

	NB	RF	LG	J48
Alle	0.639	0.61	0.674	0.606
Ersteller	0.673	0.672	0.673	0.673
Tageszeit	0.514	0.514	0.514	0.514
Enthält Medien	0.469	0.469	0.469	0.469
Ist Retweetet	0.511	0.511	0.511	0.511
Mehrfachveröffentlichungen	0.464	0.536	0.536	0.536
Anzahl Teilungen	0.563	0.518	0.566	0.543

Tabelle 6.4: ROC-Area Verlinkter Inhalt

	NB	RF	LG	J48
Alle	0.465	0.541	0.56	0.558
Zeichenlänge	0.559	0.526	0.542	0.563
Lesbarkeit	0.554	0.552	0.551	0.558
Wort- und Zeichen n-Gramme	0.571	0.642	0.645	0.557
Zeichen 1-Gramme	0.578	0.626	0.602	0.518
Zeichen 2-Gramme	0.524	0.628	0.543	0.546
Zeichen 3-Gramme	0.563	0.643	0.539	0.586
Zeichen n-Gramme	0.559	0.626	0.562	0.569
Wort 1-Gramme	0.594	0.684	0.641	0.576
Wort 2-Gramme	0.612	0.662	0.514	0.56
Wort 3-Gramme	0.561	0.607	0.509	0.562
Wort n-Gramme	0.601	0.688	0.656	0.56

Abbildung 6.1: ROC-Area der 4 Klassifikatoren



Auswertung

In Abbildung 6.1 wird dargestellt, welche ROC-Area die Klassifikatoren für die Attributgruppen erreichen konnten. In Tabelle 6.2 sind die Wahrheitsmatrizen des Random-Forest-Klassifikators der Attributgruppen dargestellt. Aus den Ergebnissen ergibt sich, dass die Tweet-Attributgruppe die beste ROC-Area im Vergleich der Gruppen erreicht. Dies bedeutet, dass eine Clickbait-Erkennung die ausschließlich die Tweets verwendet anderen Attributgruppen überlegen ist.

Die Attributgruppen *Downworthy* und *Vorgabe* erreichen nur Maximalwerte von 0.50, beziehungsweise 0.53, und sind somit Clickbait-Erkennung nicht geeignet.

Die schlechte Leistung der Attributgruppe *Downworthy* ist im seltenen Auftreten des Attributs begründet, wie schon vorher in Kapitel 5.1 beschrieben. Dies zeigt, dass der im *Downworthy*-Plugin verwendete einfache Ansatz zur allgemeinen Clickbait-Erkennung nicht geeignet ist. Die verwendeten Regexe und Ausdrücke können das Konzept Clickbait nicht besser abbilden als eine stetige Einschätzung als Nicht-Clickbait.

Die mit 14,198 Attributen größte Gruppe der Link-Attribute erreicht eine ROC-Area von etwa 0.51, also einen sich von 0.5 kaum unterscheidender Wert. Die 3748 Wort-Monogramme der *Vorgabe* erreichen dagegen bei Verwendung des Random-Forest-Klassifikators eine ROC-Area von 0.684. Es ist anzunehmen, dass der große Unterschied innerhalb der Attributgruppe durch overfitting der Klassifikatoren hervorgerufen wird. Es ist ersichtlich, dass einige

Abbildung 6.2: Wahrheitsmatrizen der Attributgruppen des Random Forest Klassifikators

Alle Attribute			Botschaft			Attributgruppe Meta		
Erkannt			Erkannt			Erkannt		
	CB	Kein-CB		CB	Kein-CB		CB	Kein-CB
CB	106	199	CB	159	146	CB	163	142
Kein-CB	67	645	Kein-CB	85	627	Kein-CB	247	465

Vorgabe			Attributgruppe Inquirer			Downworthy		
Erkannt			Erkannt			Erkannt		
	CB	Kein-CB		CB	Kein-CB		CB	Kein-CB
CB	75	230	CB	75	230	CB	6	299
Kein-CB	106	606	Kein-CB	106	606	Kein-CB	14	698

Attribute der Vorgabe durchaus geeignet sind, um Clickbait zu erkennen, aber dafür muss eine sorgfältige Attributauswahl stattfinden, da die Klassifikation sonst beeinträchtigt wird.

Die 182 Inquirer-Attribute zeigen mit einer ROC-Area von bis zu 0.659 bei dem Naiver-Bayes-Klassifikator eine im Vergleich mittlere Vorhersagekraft. Das schon in Kapitel 5.1 beschriebene Inquirer-Attribut „You“ wird von dem X^2 -Verfahren als eines der zehn entscheidendsten Attribute eingestuft.

Die Meta-Attribute zeigen eine ähnliche ROC-Area wie die Inquirer-Attribute. Das dieser Gruppe zugehörige „Ersteller“-Attribut wurde von der Attributauswahl als das wichtigste Attribut eingestuft. Das Attribut „Enthält Medien“ erzielt eine schlechtere Aussagekraft als ein Würfelwurf. Dies deutet auf Unterschiede bei der Verwendung von Medien in Trainings- und Testset hin, was dazu führt, dass sich die Klassifikatoren falsch entscheiden.

Bei Verwendung aller Attribute kann nur der Random-Forest Klassifikator eine gute Leistung erreichen.

Auch „Ist Retweet“, „Tageszeit“, „Mehrfachveröffentlichungen“ und „Teilungen“ zeigen mit Werten nur knapp über 0.5 bei allen Klassifikatoren keine dem Münzwurf überlegene Aussagekraft. Somit ist das einzige zur Clickbait-Erkennung geeignete Meta-Attribut der Ersteller, mit einer ROC-Area von 0.673.

In dieser Experimentserie hatten wir des Weiteren die Wirkung der Attributauswahl auf 10, 100 und 1000 Attribute untersucht. In dem folgenden Kapitel 6.2 wird auf dieses Thema jedoch noch tiefer eingegangen. Die Ergebnisse sind in Abbildung 6.3 dargestellt.

Dabei zeigt sich, dass die Leistung der kombinierten Tweet-Attribute bei den Klassifikatoren Naiver Bayes und Random Forest besser ist, als deren

Abbildung 6.3: ROC-Area der Klassifikatoren bei Attributauswahl

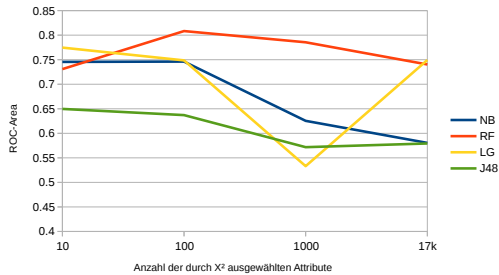


Abbildung 6.4: ROC-Kurve des RF mit 100 X^2 -ausgewählten Attributen

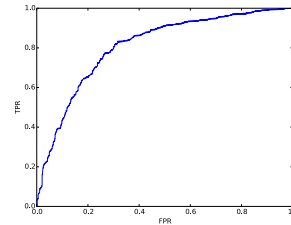


Tabelle 6.5: Wahrheitsmatrix für optimierte Konfidenz-Schwellwerte

	Erkannt	
	CB	Kein-CB
CB	143	162
Kein-CB	59	653

Bestwerte bei Attributauswahl. Dies zeigt zum einen, dass die verwendete Attributauswahl suboptimal funktioniert und dass in weiterführenden Arbeiten andere Methoden der Attributauswahl erkundet werden sollten. Zum anderen lässt sich folgern, dass sich eine Clickbait-Erkennung, welche sich auf den Tweet selbst konzentriert, gut funktioniert. Insbesondere die 543 POS-Tag-Attribute haben mit einer ROC-Area von bis zu 0.803 eine gute Erkennungsrate.

Im Vergleich der Klassifikatoren zeigt sich, dass J48 meist die schlechteste ROC-Area erreicht. Mit einer, bei Betrachtung aller Attribute und Attributgruppen, erreichten Bestleistung von 0.673 ist er deutlich schlechter als die anderen Klassifikatoren. Logistische Regression erreicht einen ROC-Area-Bestwert von 0.775 bei 10 ausgewählten Attributen, und ist somit insgesamt Drittbester Klassifikator. Der Unterschied zu Naiver Bayes ist mit 0.779 nur etwas geringer. Mit Abstand das beste Ergebnis erreichte der Random-Forest Klassifikator mit einem Bestwert von 0.814 bei Verwendung aller Tweet-Attribute.

Bei Auswertung der in Abbildung 6.4 dargestellten ROC-Area des Random-Forest-Klassifikators bei Verwendung aller Tweet-Attribute, in Verbindung mit einer Nutz-/Kosten-Analyse ergibt sich, dass sich die mit 78.2% korrekt klassifizierten Tweets beste Vorhersagengenauigkeit ergibt, falls ein Schwellwert von 0.48 für Nicht-Clickbait und entsprechend 0.52 für Clickbait angenommen wird. Die Wahrheitsmatrix für diesen Fall ist in Tabelle 6.5 dargestellt.

Attributauswahl-Kurve

Abbildung 6.5: Random Forest

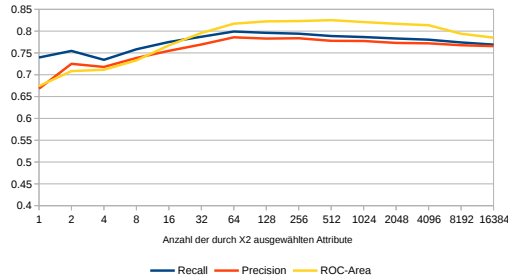
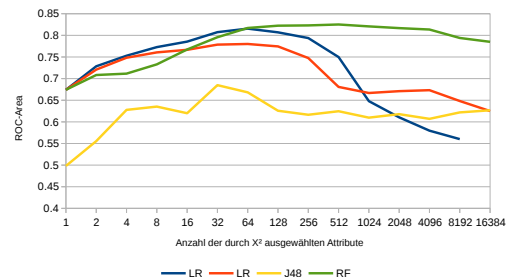


Abbildung 6.6: Vergleich Klassifikatoren



6.2 Vertiefende Experimente

Um die Experimente aus 6.1 zu vertiefen und verschiedene Fragestellungen zu untersuchen, haben wir eine Reihe weiterer Experimente durchgeführt. Um die für weitere Experimente benötigte Zeit zu reduzieren, overfitting zu vermeiden und um die Aussagekraft der Experimente sicher zu stellen, werden wir in den folgenden Experimenten meist 100 durch X^2 -Verfahren ausgewählte Attribute verwenden, da mit diesem gute Ergebnisse erzielt wurden. Wir werden Random Forest als Haupt-Klassifikator verwenden, da sich in Abschnitt 6.1 gezeigt hat, dass dieser das Modell am besten darstellen kann.

Attributauswahl-Kurve

In Abschnitt 6.1 sind große Schwankungen in der Attributauswahl aufgetreten. Daher haben wir die Leistung der Klassifikatoren bei Attributauswahl genauer untersucht. Dazu wurde eine um den Faktor 2 steigende Anzahl von Attributen mithilfe des X^2 -Verfahren ausgewählt und die Leistung der Klassifikatoren dieser Attributmengen ermittelt. Für den Random-Forest-Klassifikator haben wir Recall, Precision und ROC-Area in Abbildung 6.5 aufgeführt, in Abbildung 6.6 vergleichen wir die ROC-Areas der Klassifikator. Um die Ergebnisse besser nachvollziehen zu können, haben wir in Abbildung 6.7 die Top 8 der X^2 Attribute dargestellt. Dabei sind die Attribute in der selben Reihenfolge wie sie durch die steigendem Attributmengen eingeführt werden. *Ersteller* ist also das Attribut, welches bei nur einem ausgewähltem Attribut verwendet wird, *Ersteller* und *POS-Tag: Nomen, Singular* bei zwei Attributen und so weiter.

Für dieses Experiment wurde die Methode der 10-Fold-Cross-Validation gewählt, da die relative Leistung der Klassifikationen untereinander und in Abhängigkeit von der Menge ausgewählter Attribute im Mittelpunkt dieser

Abbildung 6.7: Die besten 8 X^2 Attribute

1. Ersteller
2. POS-Tag: Nomen, Singular
3. Länge des Tweets in Zeichen
4. Flesch-Lesbarkeit des Tweets
5. POS-Tag: Adverbien und „how“, „where“, „why“ usw.
6. Inquirer-Wortliste You
7. Verhältnis von einfachen Wörter zu allen Wörtern
8. Tweet enthält Zeichenkette: „you“

Untersuchung steht.

Auffällig ist, dass die Leistung des Random-Forest-Klassifikators bei allen Kriterien bei vier Attributen leicht abnimmt. Der Recall sinkt um 0.02 von 0.754 auf 0.735, die Precision sinkt um 0.007. Die ROC-Area hingegen steigt leicht um 0.003 auf 0.711. In diesem Bereich führen die neu eingeführten Attribute dazu, dass der Klassifikator mehr Tweets falsch bewertet. Diese Attribute scheinen ein schlechteres Modell als die vorhergehenden Attribute darzustellen.

Dabei hat Random Forest den höchsten Recall (0.800) und Precision (0.785) bei 64 Attributen erzielt, die höchste ROC-Area wurde bei 512 Attributen festgestellt (0.825), wobei hier der Unterschied zu der Area mit 64 Attributen nur 0.008 beträgt.

Auch im Vergleich der Klassifikatoren zeigt sich, dass Naiver Bayes und Random Forest den anderen überlegen sind. Bei 64 Attributen erreichen Naiver Bayes (0.816) und Logistische Regression (0.780) ihre höchste ROC-Area, J48 bei schon bei 32 Attributen (0.685). Es lässt sich beobachten, dass zu 1024 Attributen die Leistung aller Klassifikatoren stark abnimmt. Dies zeigt, dass 1024 Attribute die Obergrenze darstellt, bevor die Leistung des Modells stark abnimmt. Die weiterhin gute Leistung des Random Forest lässt sich über dessen Struktur erklären, bei welcher mit 100 Entscheidungsbäumen, die je 7 zufällige Attribute betrachten und somit die einzelnen Bäume nicht durch die große Attributmenge overfitten. Die Logistische Regression hat besonders große Unterschiede in der Leistung, der Unterschied zwischen der höchsten ROC-Area und der niedrigsten ist 0.255, praktisch die Hälfte der Skala. Dies ist ein großer Unterschied und zeigt, dass dieser Klassifikator das Modell mit einem maximal-ROC-Area von 0.816 das Modell gut darstellen kann, bei eine suboptimale Attributauswahl die Leistung dieses Klassifikator stark beeinträchtigt. Auch in diesem Experiment zeigt sich, dass J48 das Modell deutlich schlechter abbildet als die anderen Klassifikatoren.

Lern-Kurve

Abbildung 6.8: Random Forest mit 100 Attributen

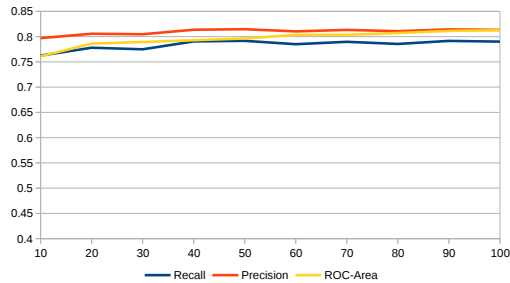
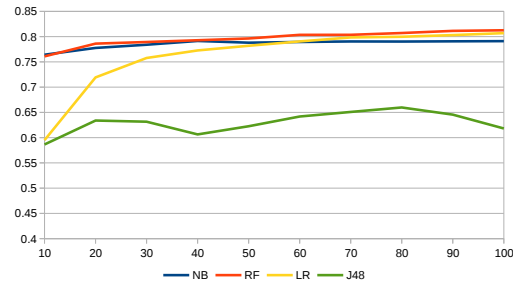


Abbildung 6.9: Klassifikatoren mit 100 Attributen



Abschließend lässt sich sagen, dass unser Modell mit 64 - 128 Attributen die beste Leistung erzielt. 100 durch das X^2 -Verfahren abgebildete Attribute sind im Anhang A.2 aufgelistet.

Lernkurve

Im Hauptexperiment wurde eine statische $\frac{2}{3}$ zu $\frac{1}{3}$ Teilung des Datensets verwendet. Um zu untersuchen, wie sich das Lernverhalten der Klassifikatoren bei unterschiedlichen Trainingssetgrößen verhält, haben wir ein Experiment durchgeführt, welches die Lernkurve der Klassifikatoren untersucht. Somit ist es möglich festzustellen, wie viele Tweets die Klassifikatoren benötigen, um das Konzept Clickbait zu erlernen.

Der folgende Vorgang wurde 10 mal durchgeführt, die Ergebnisse der einzelnen Experimente wurden gemittelt:

Mit dem X^2 -Verfahren wurden 100 Attribute ausgewählt. Daraufhin wurden zufällige 10% des Korpus als Testset festgelegt. Die übrigen 90% des Datensets wurden als Testset verwendet. Anschließend haben wir 10, 20, ..., 100% aus dem Trainingsset entnommen und als Unter-Trainingsset für den Random-Forest-Klassifikator verwendet. Die Mehrfachausführung dient dabei den Einfluss von Zufallsfaktoren, die in einzelnen Experimenten auftreten können, zu minimieren. Die Clickbait-Klassenverteilung von 4 zu 1 wurde nicht durch Oversampling korrigiert, jedoch wurde manuell überprüft, dass diese bei den Experimentwiederholungen eingehalten wurde.

Aus den in Abbildung 6.8 dargestellten Ergebnissen des Random Forest lässt sich erkennen, dass sowohl Recall als Precision nur leicht ansteigen, so steigt der Recall 0.762 auf 0.790 und die Precision von 0.777 auf 0.813. Die ROC-Area erhöht sich um 0.05 auf 0.812 und somit am meisten. Aus dem Anstieg der ROC-Area-Kurve lässt sich schließen, dass sich die Leistung des

Klassifikators mit einem größeren Trainingsset noch verbessern würde. Die sich kaum verbessernden Precision und Recall Werte sprechen dafür, dass bei einigen Tweets in unserem Korpus deren (Nicht-)Clickbait-Charakteristik noch nicht durch die von uns entwickelten Attribute abgebildet wird, da die Anzahl der *false positives* und *false negatives* hoch ist und sich deren Verhältnis zu *true positives* sich durch mehr Tweets im Trainingsset kaum erhöht.

In Abbildung 6.9 wird die ROC-Area der Klassifikatoren verglichen. Im Vergleich der Klassifikatoren zeigt sich, dass der J48-Klassifikator durch die wachsende Menge der Tweets im Trainingsset praktisch nicht profitiert. Daraus lässt sich folgern, dass dieser Klassifikator nicht zur Klassifikation des von uns erstellten Modells geeignet ist und das auch bei weiterer Vergrößerung des Korpus nicht mit einer Verbesserung der Leistung zu rechnen ist.

Der Logistische-Regression-Klassifikator verbessert Leistung durch die Zunahme des Trainingssets am meisten, es lässt sich jedoch erkennen, dass die Lernkurve einer logarithmischen Funktion gleicht und ab 70% des Testkorpus nur noch um 0.008 Punkte steigt. Somit ist davon auszugehen, dass dieser Klassifikator bei einer Vergrößerung des Korpus nicht mit der Verbesserung der ROC-Area profitiert.

Sowohl Naiver Bayes und Random Forest erlernen das Konzept Clickbait sehr schnell und erreichen ab 10% des Trainingssets eine ROC-Area von 0.76. Bei Vergrößerung des Trainingssets verbessert sich Naive Bayes, die ROC-Area jedoch nur um 0.027, Random Forest hingegen um 0.051.

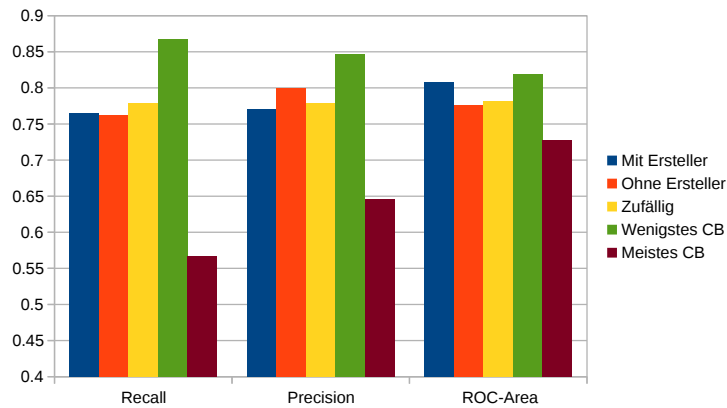
Bei keinem Klassifikator steigt die Leistung nach etwa 60% des Trainingssets noch relevant an. Dies lässt darauf schließen, dass eine Vergrößerung des Korpus bei Beibehaltung des aktuellen Modells wahrscheinlich keine größere Verbesserung der Leistung hervorrufen würde.

Einfluss des Erstellers

Da wie zuvor in Abschnitt 5.2 diskutiert die Verwendung des Erstellers eines Tweet, beziehungsweise der Herausgeber, als Attribut nicht selbstverständlich ist, soll im Folgenden der Einfluss dieses Attributs analysiert werden. Die folgenden Experimente ermöglichen es weiterhin eine Abschätzung zu treffen, wie sich das Modell verhält, falls es mit neuen Herausgebern konfrontiert wird, die einen anderen Stil, oder andere Clickbait- zu Nicht-Clickbait-Verhältnisse veröffentlichen. Die Ergebnisse dieser Experimente sind in Abbildung 6.10 dargestellt. Als Klassifikator verwenden wir in allen Experimenten in diesem Abschnitt Random Forest. Die Vergleichswerte, *Mit Ersteller*, sind aus dem Hauptexperiment übernommen, mit 100 durch X^2 ausgewählten Attributen und enthalten das Attribut Ersteller.

In einem ersten Experiment testen wir die Leistung unseres Modells oh-

Abbildung 6.10: Vergleich der Leistung abhängig vom Ersteller



ne das Attribut Ersteller. Dazu wird das Datenset wie zuvor mit einem $\frac{2}{3}$ Trainingsset zu $\frac{1}{3}$ Testset geteilt. Das Attribut Ersteller wurde entfernt und daraufhin wurden mit dem X^2 -Verfahren 100 Attribute ausgewählt. Dieses Experiment dient zum einen dazu, den Einfluss des Attributs auf die Klassifikatorleistung zu messen. Zum anderen ist es notwendig, um zu gewährleisten, dass die kommenden Experimente mit dem Hauptexperiment vergleichbar bleiben.

In einem zweiten Experiment haben wir die Tweets von fünf zufälligen Herausgebern als Testset und die Tweets der anderen 15 als Trainingsset verwendet. Dabei wurde ebenfalls das Attribut *Ersteller* entfernt, da es aufgrund der Ausschließlichkeit der Ersteller keine Information enthalten kann. Dieses Experiment wurde zehn mal durchgeführt und aus den jeweils entstanden Werten der Mittelwert gebildet. Bei diesem Experiment sollte insbesondere untersucht werden, wie groß der Einfluss des Herausgeberstils ist.

In einem dritten und vierten Experiment haben wir jeweils die fünf Herausgeber mit dem meisten beziehungsweise dem wenigsten veröffentlichtem Clickbait als Trainingsset verwendet. Diese Experimente ermöglichen es uns, eine Aussage darüber treffen zu können, wie sich das Modell verhält, falls die Tweets eines anderen Herausgebers bewertet werden der deutlich mehr oder deutlich weniger Clickbait veröffentlicht.

Aus den Ergebnissen lässt sich erkennen, dass sich das Attribut Ersteller in Recall und ROC-Area positiv auf die Leistung auswirkt, die Precision jedoch ohne Ersteller höher ist. Es folgt daraus, dass das Entfernen des Erstellers die Anzahl der *false positives* verringert, was darauf schließen lässt, dass das Attribut Ersteller einen Bias zu einer Klasse aufgrund des Erstellers einführt. Der Grund dafür ist, dass bei einem Ersteller der sehr viel Clickbait veröffentlicht, zu erwarten ist, dass weitere Tweets von ihm ebenfalls Clickbait sind.

Dies ruft jedoch auch eine Voreingenommenheit des Klassifikator hervor. Dies wird durch die Ergebnisse der Experimente mit zufälligem Ersteller im Trainingsset deutlich, da bei diesen Recall und Precision durchschnittlich bessere Werte erreichen als bei Vorhandensein des Erstellers. Jedoch ist auch hier die ROC-Area mit Ersteller besser.

Falls im Testset die Herausgeber mit dem niedrigsten Anteil an Clickbait sind, erreicht der Klassifikator in allen Bereichen die beste Leistung, bei denen mit dem größten Anteil die mit Abstand schlechteste. Dies zeigt, dass unser Modell besser für Herausgeber mit wenig Clickbait funktioniert und deutet darauf hin, dass unser Modell besser darin ist Nicht-Clickbait zu erkennen, als Clickbait selber. Die Ergebnisse deuten auch daraufhin, dass das Attribut Ersteller eher einen Bias gegenüber Tweets eines Erstellers erzeugt und dass es für eine faire Beurteilung des Stils eines Tweets ratsam ist, dieses Attribut nicht zu verwenden oder geringer zu gewichten.

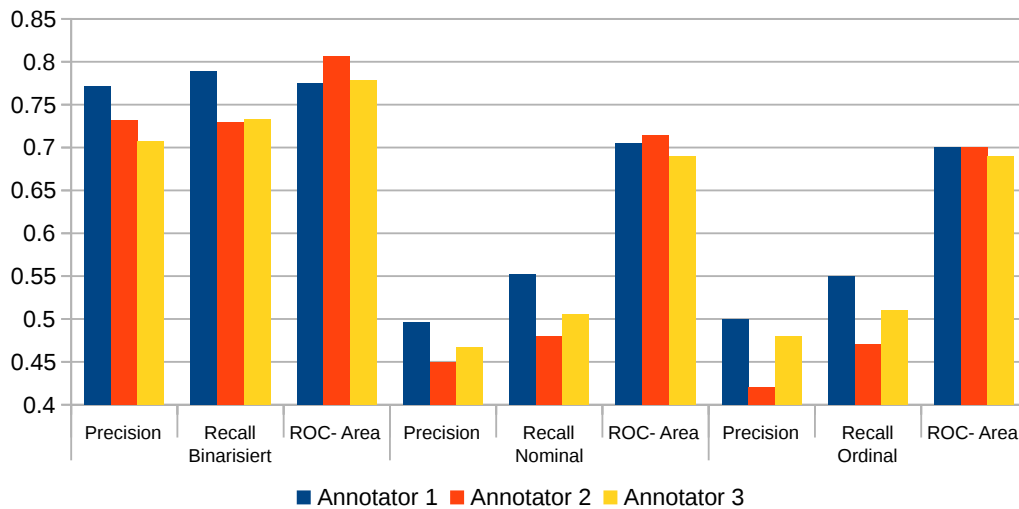
Es lässt sich schlussfolgern, dass unser Modell bei Einführung von neuen Erstellern, oder neuen Autoren, die einen anderen Stil haben oder deutlich mehr Clickbait veröffentlichen, wahrscheinlich keine mit den Ergebnissen aus 6.1 vergleichbare Leistung erbringen kann. Abschließend lässt sich sagen, dass das Attribut Ersteller nicht essentiell ist und dass es die Leistung des Klassifikators in Verbindung mit den in Kapitel 5.2 genannten Argumenten rechtfertigen, dieses Attribut nicht zu verwenden.

Einfluss der Reihenfolge

Im Hauptexperiment wurde streng darauf geachtet, dem Klassifikator keine Informationen aus der *Zukunft* der untersuchten Woche, die das Trainingsset darstellt zukommen zu lassen. Um nachvollziehen zu können, wie groß der Einfluss des zeitlichen Auftretens der Tweets ist, wurde ein Experiment durchgeführt, bei welchem vor der Teilung des Datensets in Trainings- und Testset mit dem X^2 -Verfahren 100 Attribute ausgewählt wurden. Dies bedeutet, dass die ausgewählten Attribute im gesamten Datenset eine hohe statistische Abhängigkeit mit der Klasse Clickbait haben. Daraufhin wurde die Reihenfolge der Tweets im Datenset randomisiert. Danach wurde das Datenset wie zuvor in Trainings- und Testset geteilt und das Trainingsset oversampled. Es wurde der Random-Forest-Klassifikator verwendet.

Bei Durchführung dieses Experiments wurde ein Recall von 0.785, eine Precision von 0.773 und eine ROC-Area von 0.804 festgestellt. Diese Werte entsprechen weitgehend den Ergebnissen aus 6.1, woraus sich schließen lässt, dass der zeitliche Einfluss gering ist. Dies bestätigen die 100 ausgewählten Attribute, bei denen weder im Hauptexperiment noch in diesem Experiment Namen, Personen oder Orte enthalten sind.

Abbildung 6.11: Ergebnisse mit annotatorenspezifischem Modell



Daraus folgt, dass unser Modell auch neue Tweets, die tagesaktuelle oder andere dem Modell *unbekannte* Themen behandeln, mit vergleichbarer Leistung funktioniert.

Einzelannotatoren

In den folgenden Experimentserien soll untersucht werden, wie gut Clickbait-Erkennung für einzelne Personen funktioniert. Dadurch soll die in Kapitel 1 gestellte Frage „Ist es praktikabel, Clickbait-Erkennung zu personalisieren, also für die Bedürfnisse einer einzelnen Person anzupassen?“ beantwortet werden. Die Ergebnisse dieser Experimente werden in Abbildung 6.11 zusammen gefasst.

Die Motivation dafür, wie zuvor genauer in Kapitel 4 beschrieben, ist der Umstand, dass die Annotatoren bemerkbare in der Unterscheidung von Clickbait hatten. Durch diese Experimente soll festgestellt werden, ob es dem Klassifikator möglich ist, persönliche Schwerpunkte in der Wahrnehmung von Clickbait korrekt festzustellen.

Dabei ist die grundsätzliche Vorgehensweise bei den drei Experimenten gleich. Das Datenset mit den Clickbait-Klassen des jeweiligen Annotators wurde daraufhin mittels $\frac{2}{3} - \frac{1}{3}$ -Splits in ein Trainings- und Testset geteilt und 100 Attribute mittels des X^2 -Verfahrens selektiert. Daraufhin wurde das Trainingsset oversampled. Als Klassifikator wurde Random Forest verwendet.

Binarisiert

Zunächst haben wir zuvor die Annotation jedes Annotators balanciert binarisiert. Aus den Ergebnissen lässt sich erkennen, dass zwischen den Annotatoren nur kleine Unterschiede gibt und die erzielten Werte denen im Hauptexperiment sehr ähnlich sind. Dies zeigt zum einen, dass eine persönliche Clickbait-Erkennung mit unserem Modell genauso gut funktioniert wie eine allgemeine. Jedoch bietet diese momentan keine Vorteile, im Sinne einer verbesserten Leistung, gegenüber dem allgemeinen Clickbait-Modell.

Nicht Binarisiert

In einem weiterführenden Experiment haben wir die annotierten Klassen Nicht-, Schwacher-, Mittlerer- und Starker-Clickbait verwendet. Hierbei lässt sich erkennen, dass Erkennung deutlich schlechter funktioniert. Precision und Recall haben Werte um 0.5, dies ist sehr schlecht, da es einem Würfelwurf entspricht. Die ROC-Area ist mit etwa 0.7 bei allen Annotatoren um 0.1 schlechter als bei Verwendung der binarisierten Werte. Die Erkennung der 4 ursprünglich annotierten Klassen funktioniert mit unserem Modell deutlich schlechter als bei einer binarisierten Betrachtung. Dies lässt sich durch die gestiegene Komplexität der Klassifikationsaufgabe durch die Erhöhung der zu erkennenden Klassen begründen. Es ist jedoch auch anzumerken, dass die Klassen untereinander deutlich ungenauer abgegrenzt sind, da viele Merkmale zusammenspielen um Clickbait zu definieren.

Ordinal nicht binarisiert

Da die annotierten Klassen Nicht-, Schwacher-, Mittlerer- und Starker Clickbait ordinal sind, also natürliche Ordnung besitzen, haben wir ein Experiment erstellt, welches einen ordinalen Klassifikator benutzt. Dieser ist in dem von uns verwendeten Weka-Framework in Form eines Meta-Klassifikator gegeben[55]. Die entstanden Werte sind den Ergebnissen der vorhergegangenen Experimente sehr ähnlich. Somit kann durch die verwendete ordinale Klassifikation keine Verbesserung der Erkennung gezogen werden.

6.3 Manuelle Attributmenge

Anhand der aus den vorherigen Experimenten gewonnen Erkenntnissen werden wir im Folgenden ein Attributset selbst erstellen und dessen Leistung testen. Dabei wurde eine ROC-Area von 0.826, eine Precision von 0.784 und ein Recall von 0.798 erreicht.

Es werden Attribute aus den in Abschnitt 6.1 gut funktionierenden Attributen und Attributgruppen verwendet. Aufgrund der guten Leistung des Modells, auch ohne Ersteller, werden wir auf das Ersteller-Attribut verzichten. In Abschnitt 6.2 konnte gezeigt werden, dass das Modell mit 64-128 Attributen die beste Leistung erreicht, daher streben wir eine ähnliche Anzahl an verwendeten Attributen an. Da in Abschnitt 6.2 festgestellt wurde, dass das Modell nicht abhängig von der gewählten Reihenfolge der Tweets ist, verwenden wir 10-Fold-Cross-Validation, um das Experiment auszuwerten. Als Klassifikator wurde wie zuvor Random Forest verwendet.

Die erstellte Attributmenge mit 92 Attributen enthält 10 POS-Tag-, 47 Tweet-BoW- und 30 verlinkte-Inhalte-Bow-Attribute. Weiterhin enthalten sind die Attribute Tweet Zeichenlänge, Tweet startet mit Zahl, Tweet-Lesbarkeit, Verhältnis einfacher Wörter zu allen Wörtern und Anzahl der Präpositionen.

Der vorherige ROC-Area-Bestwert von 0.825 bei 512 durch X^2 ausgewählten Attributen konnte in diesem Experiment nicht nennenswert überboten werden, auch Recall und Precision konnten nicht erhöht werden. Jedoch benötigt die im Anhang A.3 gezeigt Attributmenge nicht das Ersteller-Attribut und nur ein fünftel der Attribute.

6.4 Fehleranalyse

Im Folgenden soll untersucht werden, welche Tweets bei unserer Hauptexperimentreihe in Kapitel 6.1 der Random-Forest-Klassifikator mit 100 durch X^2 ausgewählten Attributen falsch bewertet hat, um daraus zu folgern, wie das Modell weiter verbessert werden kann, um eine bessere Leistung zu erreichen.

Stichprobenanalyse

Zunächst haben wir eine Stichprobenanalyse durchgeführt, um Fehler in dem Modell zu suchen. Dabei hat sich ergeben, dass das entstandene Modell noch nicht ausreichend in der Lage ist, dass für Clickbait typische Fehlen essentieller Information in komplexeren Tweets zu erkennen. So wurden folgende Clickbait-Tweets falsch klassifiziert:

- The Sun criticises Newcastle's reporter blacklist with special back page
- A planned 1,340-foot building at Two World Trade Center will honor the old and the new
- What D.C. posts on Twitter versus what everybody else posts on Twitter, visualized

Diese Tweets haben gemeinsam, dass sie nicht die für Clickbait typischen Formulierungen verwenden, aber dennoch subtil für das Verständnis wichtige Informationen ausklammern. Bei einigen Tweets scheint das Trainingsset noch nicht ausreichend divers zu sein, um die Tweets als Clickbait zu erkennen.

- There's one thing Wall Street interns care about more than which bank they get into
- 13 of America's best restaurants — in gas stations

Eine Vergrößerung des „Wortschatzes“ der Attribute durch eine Erweiterung des Korpus würde diesen Fehler entgegenwirken. Dies widerspricht den Ergebnissen aus Abschnitt 6.2, bei welchen sich abgezeichnet hat, dass sich eine Korpusvergrößerung nicht positiv auf die Leistung der Klassifizierer auswirkt. Es scheint jedoch schlüssig, dass eine Vergrößerung des Korpus diese Fehler beheben würde.

Konfidenz

Ein Klassifikator gibt bei der Bewertung von Konzepten die sogenannte Konfidenz an, ein Wert der beschreibt, wie sicher dieser sich bei der Bestimmung der Klasse ist. Wir wollen diese im Folgenden im Allgemeinen und im Hinblick auf einige Kriterien betrachten, um festzustellen ob sich dadurch auf Probleme in unserem Modell schließen lässt. In dem Boxplot von Abbildung 6.12 ist zu sehen, wie sich die Konfidenz des Klassifikators verhält, in Bezug darauf, ob er einen Tweet richtig eingeschätzt hat. Daraus geht hervor, dass der Klassifikator bei Tweets, bei welchen er eine hohe Konfidenz hat, öfter richtig liegt. Dies zeigt, dass sich durch eine geeignete Konfidenzschwelle einige falsche Bewertungen bezüglich der Klasse Clickbait (false positives) vermeiden lassen. Um diese Untersuchung weiter zu differenzieren, haben wir diese Untersuchung für die annotierte Clickbait-Klasse wiederholt, wie in Abbildung 6.13 dargestellt. Dabei fällt auf, dass bei Clickbait-Tweets die zweite und dritte Quantile der Konfidenz mit einem Median der Konfidenz von etwa 0.6 niedrig ist. Die Konfidenz bei der Bewertung von Clickbait ist sehr niedrig und nur unwesentlich höher bei korrekter Bewertung. Dies wird durch 6.14 Nicht-Clickbait unterstrichen, bei welchem sich erkennen lässt, dass der Klassifikator hier bei richtigen Antworten eine deutlich höhere Konfidenz hat. Die deutlich höhere Konfidenz bei der Einschätzung von Nicht-Clickbait lässt darauf schließen, dass der Klassifikator besser darin ist Nicht-Clickbait zu erkennen. Es scheint leichter zu sein, Nicht-Clickbait zu erkennen, als Clickbait selbst. Da wir einige Attribute entwickelt haben, die es dem Klassifikator erlauben sollten Clickbait zu erkennen, etwa *Beginnt mit Zahl*, *Anzahl Punkte* u.a., deutet die Unsicherheit bei Clickbait darauf hin, dass diese Attribute noch nicht in der Lage sind

Vergleich der Klassifikatorkonfidenz

Abbildung 6.12: Klassenunabhängig

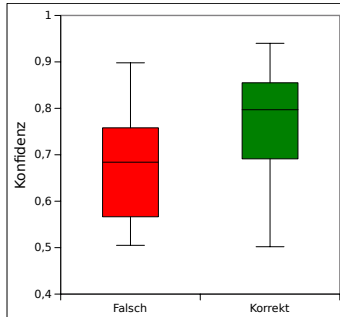


Abbildung 6.13: Clickbait

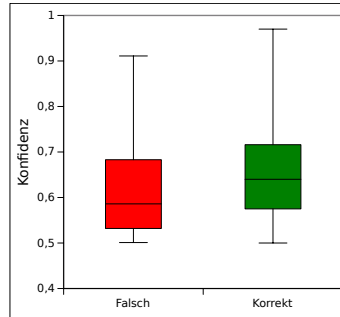


Abbildung 6.14: Nicht-Clickbait

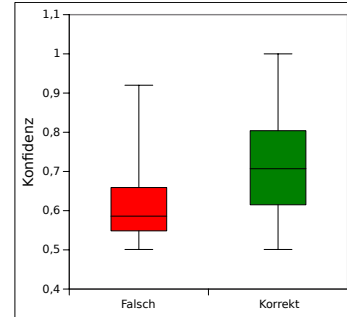
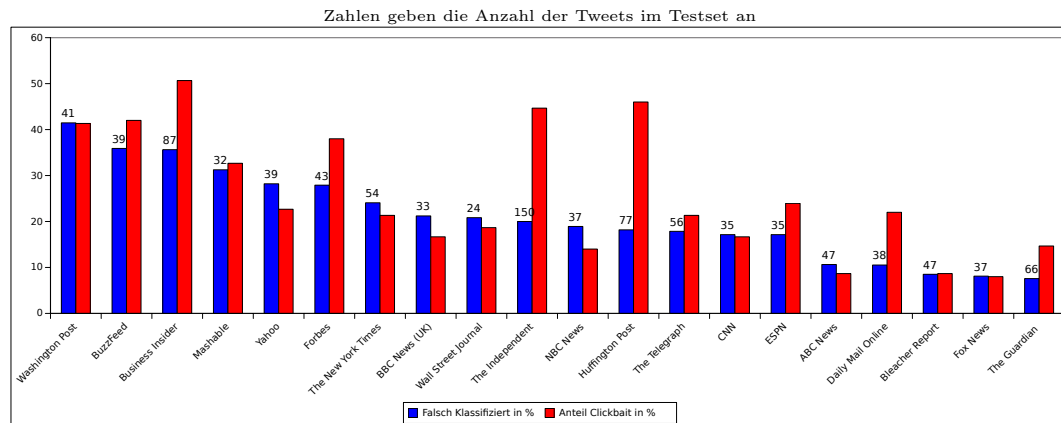


Abbildung 6.15: Anteil der falschen Vorhersagen des Klassifikators nach Ersteller



Clickbait verlässlich zu erkennen. Eine andere Erklärung ist hingegen, dass Clickbait in unserem Korpus so vielfältig erscheint, dass es dem Klassifikator schwer fällt das Konzept selbst zu erlernen und dass es stattdessen leichter ist Nicht-Clickbait zu trainieren.

Abhängigkeit vom Ersteller

Da der Ersteller eines Tweet in unserem Modell eine große Rolle innehält, wie man an der Priorität des Attributes bei der X^2 -Attributauswahl sieht, haben wir dessen Einfluss genauer untersucht. In Abbildung 6.15 haben wir daher dargestellt, welchen Anteil der Tweets bei den Nutzern falsch bewertet wurde, unabhängig davon ob ein Tweet Clickbait ist oder nicht und wie viele der Tweets des Herausgebers Clickbait sind. Wie zu sehen ist, entspricht der Anteil

Vergleich der Vorhersagen

Abbildung 6.16: Nach Medien

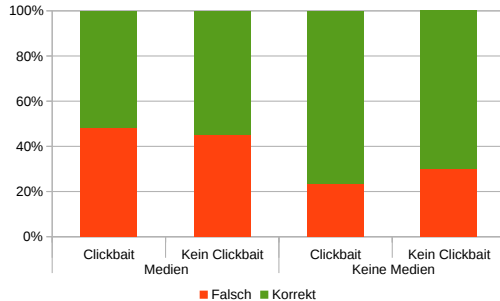
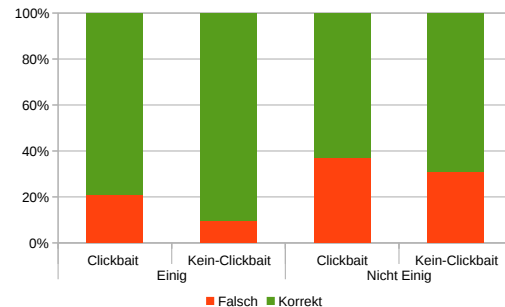


Abbildung 6.17: Nach Annotatoreneinig-
nigkeit



der falsch bewerteten Tweets dem Anteil an veröffentlichten Clickbait. Hervorstechend sind die Herausgeber „The Independent“ und „Huffington Post“, bei welchen der Klassifikator deutlich besser in der Lage ist, Clickbait zu erkennen. *The Independence* ist ein besonderer Fall, da dessen 150 Tweets ausschließlich im Testset sind. Obwohl der Klassifikator in Trainingsphase mit keinen Tweets dieses Erstellers arbeiten konnte, wurden hier nur 20% der Tweets falsch bewertet. Dies spricht dafür, dass dieser Herausgeber ein mit den Modell des Klassifikators gut vereinbarenden Stil hat.

Abhängigkeit von Medien

Da bereits festgestellt wurde, dass die Existenz von Medien und deren Inhalte eine große Rolle für Clickbait spielen, haben wir deren Einfluss auf die Leistung des Klassifikators in Abbildung 6.16 dargestellt. Es lässt sich erkennen, dass Medien die Leistung stark beeinflussen, so ist bei Clickbait die Wahrscheinlichkeit, dass der Klassifikator einen Tweet falsch einschätzt etwa doppelt so hoch. Dies zeigt, dass es deutlich schwieriger ist, Clickbait zu erkennen falls ein Tweet Medien enthält, da diese wie schon in Kapitel 3 gezeigt, großen Einfluss auf die Clickbait-Klasse haben können.

Abhängigkeit von Annotatoreneinverständnis

Da die Clickbait-Klasse durch die Kombination mehrerer Annotatorenmeinungen entstanden ist, haben wir die Leistung des Klassifikators abhängig davon untersucht, ob sich die Annotatoren darüber einig bezüglich der Clickbait-Klasse sind. Dazu haben die Einzelannotationen binarisiert und falls alle Annotatoren der Meinung waren, dass ein Tweet Clickbait oder Nicht-Clickbait

ist, wurde dies als *Einig* gewertet. Die Ergebnisse dieser Analyse sind in Abbildung 6.17 dargestellt. Hierbei konnten wir feststellen, dass der Klassifikator Tweets doppelt so oft falsch einschätzen, wenn sich die Annotatoren nicht einig sind. Dies zeigt, dass Tweets, bei denen sich Menschen nicht einig sind, auch schwerer für den Klassifikator zu bestimmen sind. Da die Nichteinigkeit oftmals durch die Anwesenheit von Clickbait-Merkmalen hervorgerufen wird, welche jedoch von den Annotatoren unterschiedlich stark bewertet werden, spiegelt sich dies in den Attributen der Tweet wieder. Dies macht die Einschätzung für den Klassifikator schwerer.

Abschließend lässt sich sagen, dass Tweets, insbesondere Nicht-Clickbait Tweets, die für Menschen eindeutig zu bestimmen sind, auch für den Klassifikator leichter zu bestimmen sind.

Kapitel 7

Webservice

In einem auf den Ergebnissen der Experimente aufbauenden Schritt haben wir eine zur Demonstration unserer Ergebnisse bestimmten Webservice entwickelt. Dieser besteht aus zwei Teilen. Die Grundlage bildet eine Server-API die es ermöglicht Tweets und einfache Texte durch einen Klassifikator in Clickbait und Nicht-Clickbait zu kategorisieren. Weiterhin wurde eine Client-Demo in Form einer Internetseite entwickelt, welche die vom Server gelieferten Bewertungen verwendet, um die letzten Tweets der von uns untersuchten Herausgeber je nach Klassifikation farblich gekennzeichnet darzustellen.

Im Folgenden werden wir auf die Motivation für dessen Erstellung, Aufbau und Funktionen eingehen.

7.1 Motivation

Ziel des Webservices ist es zum einen, unsere Ergebnisse zu demonstrieren, aber auch einen Grundstein zu legen, mit welchen wir die Ergebnisse unserer Arbeit weiter testen und der Öffentlichkeit zugänglich machen können.

Ziel war es weiterhin, einen ersten Bezugspunkt zu schaffen, mit welchem die Entwickler von Browser-Plugins, Webdiensten und Anderem eine Möglichkeit haben, die von uns gewonnen Erkenntnisse in automatischer Clickbait-Erkennung in ihre Arbeit einzubeziehen.

7.2 Umsetzung

Der Webservice basiert auf einer Server-API, die die Klassifikation von Tweets und Texten erlaubt. Die direkte Klassifikation von Tweets erlaubt es die Bandbreite an von uns entwickelten Attributen einzusetzen. Die Bewertung einfacher Texte soll es erlauben anderweitig gewonnene Überschriften zu bewerten.

Dazu haben wir einen Java Servlet implementiert, welches auf einem Tomcat Server eingesetzt wird. Für die Clientseite haben wir Kachellayout gewählt, welches mit den JavaScript-Bibliotheken Bootstrap und Masonry umgesetzt wurde. Um die verschiedenen von uns entworfenen Modellkonfigurationen, beispielsweise bei Verwendung von ausschließlich Tweetsattributen oder 100 durch X^2 ausgewählte Attribute, praktisch vergleichen zu können, ist es möglich mehrere Konfigurationen zu laden und über die API anzufordern.

Server

Bei starten des Servers lädt dieser die vorgegebenen Modelle um daraufhin auf Anfragen zu warten. Die serverseitige API erwartet GET-Anfragen, dabei akzeptiert sie Tweets, spezifiziert über ihre Twitter-ID und einfache Tweettexte. Weiterhin können Twitternutzer erfragt werden, in diesem Fall werden die letzten 20 Tweets des Nutzers kategorisiert. Die Antwort des Servers erfolgt in Form von JSON-Daten. Dies JSON-Daten enthalten die vorhergesagte Klasse, die Konfidenzverteilung des Klassifikators, das für die Klassifizierung verwendete Modell, sowie den bewerteten Text.

Client

In Abbildung 7.1 ist der zur Demonstration bestimmte Webservice dargestellt. Er ermöglicht es einen der 20 von uns untersuchten Herausgeber auszuwählen, woraufhin dessen 20 letzte Tweets dargestellt werden. Für die Darstellung wird ein Kacheldesign verwendet, da dies die durch Tweetlänge und Medien stark variierenden Größen der Tweets platzsparend und übersichtlich abbilden kann. Zusätzlich zu Text und Medien des Tweets wird ein Link auf die Twitterseite des Tweets angeboten und der Zeitpunkt der Tweeterstellung angegeben.


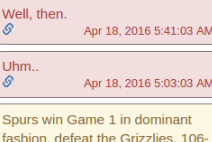
Die gesamte Webseite wurde mit Bootstrap implementiert. Für das Kacheldesign wurde Masonry, eine JavaScript-Bibliothek für dynamische Grid-Designs verwendet.

Die durch das Modell ermittelte Klasse wird farblich kodiert dargestellt. Ein roter Hintergrund deutet an, dass ein Tweet als Clickbait eingeschätzt wird, während ein grüner Hintergrund eine Einschätzung als Nicht-Clickbait anzeigt. Falls der Klassifikator eine Konfidenz von weniger als 0.6 hat, sich also in seiner Klassifikation unsicher ist, wird dies durch einen gelben Hintergrund angezeigt. Dies ermöglicht das schnelle erkennen ob ein Tweet als Clickbait identifiziert wurde. Der gelbe Hintergrund für eine unsichere Klassifikation signalisiert, dass für diesen Tweet keine ausreichend genaue Bewertung statt finden kann. Somit wird die Anzahl falscher Einstufungen deutlich reduziert.

Abbildung 7.1: Die Demo

Demo

Bleacher Report ▾

<p>? Apr 18, 2016 7:18:40 AM</p> <p>At the end of the first half, the Clippers lead the Blazers, 50-42 #NBAPlayoffs</p>  <p>Apr 18, 2016 5:45:46 AM</p>	<p>The Clippers take Game 1 against the Blazers with a 115-95 victory #NBAPlayoffs</p>  <p>Apr 18, 2016 7:17:54 AM</p>	<p>Hack-A-Jordan got me like...</p>  <p>Apr 18, 2016 7:07:15 AM</p>	<p>Looks like Mason Plumlee still hasn't noticed..BLAKE is BACK</p>  <p>Apr 18, 2016 6:22:33 AM</p>
<p>Dame Dolla vs. CP3: Battle of the point guards, as the Blazers and Clippers face-off in the First Round.</p>  <p>Apr 18, 2016 3:03:51 AM</p>	<p>That Kawhi Leonard defense can bring out all the frustration</p>  <p>Apr 18, 2016 3:24:42 AM</p>	<p>BLAKE is BACK</p>  <p>Apr 18, 2016 4:54:52 AM</p>	<p>Well, then.</p>  <p>Apr 18, 2016 5:41:03 AM</p>
<p>RT @BR CFB: Class of 2017 WR</p>	<p>Luke Walton reportedly indicated to Phil Jackson that he is not interested in the Knicks job</p>  <p>Apr 18, 2016 3:57:25 AM</p>	<p>After 3 quarters: Spurs - 81 Grizzlies - 51 #NBAPlayoffs</p>  <p>Apr 18, 2016 3:57:25 AM</p>	<p>Uhm..</p>  <p>Apr 18, 2016 5:03:03 AM</p>
			<p>Spurs win Game 1 in dominant fashion, defeat the Grizzlies, 106-74 #NBAPlayoffs</p> 

Kapitel 8

Zusammenfassung und Ausblick

In diesem Kapitel werden wir die Ergebnisse dieser Arbeit zusammenfassen, um daraus auf die sich aus diesen Beobachtungen und Experimenten entwickelnden neuen Fragestellungen und zukünftigen Forschungsbereiche zum Bereich automatische Clickbait-Erkennung einzugehen.

Wir haben in dieser Arbeit eine umfassende Übersicht über Merkmale und Besonderheiten von Clickbait erstellt, mit welchen dieser identifiziert werden kann, und somit eine differenzierte Bewertung Clickbaits erlaubt.

Bei Annotation des Korpus und der Analyse dieser haben wir erkannt, dass es trotz vorheriger Besprechungen deutliche Unterschiede in der Wahrnehmung von Clickbait und der Priorisierung von dessen Merkmalen gibt. Im Vergleich der Annotation mehrfach erscheinender Botschaften ergaben sich bemerkbare Unterschiede bei wiederholter Betrachtung. Beides unterstreicht, dass Clickbait ein schwer zu greifendes Thema ist zu dessen Verständnis weitere Experimente nötig sind, aber auch die Notwendigkeit eine größeres Korpus zu erstellen und durch mehr Annotatoren ein klareres und konsistenteres Bild von Clickbait zu schaffen.

Für zukünftige Arbeiten sollten spezielle Ausprägungen von Clickbait, beispielsweise Quizze, Listicles und Videos tiefer analysiert werden, um eine spezielle Erkennung zu ermöglichen.

In dem Hauptexperiment in Abschnitt 6.1 konnte gezeigt werden, dass es uns zu diesem Zeitpunkt mit, einer ROC-Area von etwa 0.826, guter Leistung möglich ist, Clickbait zu erkennen. Dabei hat sich gezeigt, dass es leichter ist Nicht-Clickbait festzustellen, als Clickbait. Das Konzept Clickbait konnte noch nicht in allen Bereichen ausreichend erfasst werden, insbesondere Medien und Grenzwerte, zu denen die Annotatoren unterschiedlicher Meinung waren, haben sich als Herausforderung für den Klassifikator herausgestellt.

Um das Korpus in Hinsicht auf die Clickbait-Stufen ausgewogener zu machen und um die in die Kapitel 6.4 festgestellte Unsicherheit des Klassifikators

bei der Bestimmung von Clickbait zu reduzieren, sollten bei zukünftiger Erweiterung des Korpus bewusst Herausgeber verwendet werden, die mehr Clickbait verbreiten.

Wir konnten zeigen, dass es möglich ist Clickbait ohne Leistungseinbußen ausschließlich anhand des Tweettextes und der zugehörigen Medien zu identifizieren. Des Weiteren konnten wir zeigen, dass das Modell bei Attributauswahl, um optimale Leistung zu erzielen, keinerlei Themen-, Personen- oder Ortsabhängigkeit besitzt und wir somit in dieser Hinsicht wie von uns angestrebt universell ist. Vielmehr basiert es, durch die gute Leistung der NLP-Attribute, typischer Worte wie „you“, „This“ und „what“ und der Wortlisten, auf der Analyse des Stils einer Botschaft und des dazugehörigen Links. Aus den Beobachtungen in Abschnitt 6.4 geht hervor, dass das Modell noch nicht in der Lage ist, die in Kapitel 4.5 beschriebenen komplexen Zusammenhänge zwischen Medien und Clickbait darzustellen. Daraus geht hervor, dass die Vertiefung dieser Beziehungen eine für zukünftige Arbeiten unerlässliche Aufgabe ist.

Es war uns möglich zu zeigen, dass der Ersteller einer Botschaft als Attribut zur Gewährleistung der Leistungsfähigkeit nicht nötig ist, was es uns ermöglicht Botschaft unvoreingenommen bezüglich des Herausgebers zu erkennen. Jedoch haben unsere Experimente gezeigt, dass das Modell bei der Bewertung eines neuen Herausgebers mit unbekanntem Clickbait-Anteil deutlich ungenauer werden kann.

Es ist für die binarisierte Betrachtung Clickbaits mit einer der des Hauptexperiments entsprechenden Leistung möglich auf die Präferenzen und persönlichen Schwerpunkte der Annotatoren einzugehen. Die einzelnen Klassen sind zum jetzigen Zeitpunkt nominal und ordinal nicht eindeutig erkennbar, eine genaue Abgrenzung zwischen den Klassen ließ sich noch nicht feststellen. So bleibt für zukünftige Forschung die Aufgabe, Möglichkeiten zu finden die Grenzen zwischen den Stufen genau zu identifizieren.

Der implementierte Webservice ermöglicht die angestrebte automatische Clickbait-Erkennung auf Grundlage der von uns gewonnen Erkenntnisse. Schon jetzt ist anhand des Demo-Clients erkennbar wie vorteilhaft sich eine Clickbait-Erkennung den Konsum journalistischer Inhalte auswirken kann.

Literaturverzeichnis

- [1] Martin Potthast und Sebastian Köpsel und Benno Stein und Matthias Hagen. In *Advances in Information Retrieval. 38th European Conference on IR Research (ECIR 16)*, Berlin Heidelberg New York. 1, 2.4
- [2] Daniel Dor. On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5):695–721, 2003. 2.1
- [3] Jonas Nygaard Blom and Kenneth Reinecke Hansen. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76: 87–100, 2015. 2.1, 2.3, 3.1
- [4] Thomas Patterson. *Doing well and doing good: How soft news and critical journalism are shrinking the news audience and weakening democracy-and what news outlets can do about it*. Joan Shorenstein Center on the Press, Politics and Public Policy, John F. Kennedy School of Government, Harvard University, 2000. 2.1
- [5] Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as? false news? In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, pages 15–19. ACM, 2015. 2.2
- [6] Khalid El-Arini und Joyce Tang. News feed fyi: Click-baiting. URL <http://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting/>. Zuletzt Besucht: 2016-01-25. 2.2, 4.1
- [7] Elly Ifantidou. Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics*, 41(4):699–720, 2009. 2.4
- [8] Jonah Berger and Katherine L Milkman. What makes online content viral? *Journal of marketing research*, 49(2):192–205, 2012. 2.4
- [9] Marco Guerini and Jacopo Staiano. Deep feelings: A massive cross-lingual study on the relation between emotions and virality. In *Proceedings of*

- the 24th International Conference on World Wide Web Companion*, pages 299–305. International World Wide Web Conferences Steering Committee, 2015. 2.4, 3.1
- [10] Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(03):715–734, 2005. 2.4
 - [11] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207, 2013. 2.4
 - [12] Elena Hensinger, Ilias Flaounas, and Nello Cristianini. Modelling and explaining online news preferences. In *Pattern Recognition-Applications and Methods*, pages 65–77. Springer, 2013. 2.4
 - [13] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! *Icwsn*, 11:538–541, 2011. 2.4
 - [14] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014. 2.4, 5.1, 5.1
 - [15] Andrew K Przybylski, Kou Murayama, Cody R DeHaan, and Valerie Gladwell. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior*, 29(4):1841–1848, 2013. 2.4
 - [16] Nadya Khoja. Capitalizing on fomo: Why the fear of missing out is great for businesses, July 2015. URL <http://rdmr.ca/capitalizing-on-fomo-why-the-fear-of-missing-out-is-great-for-businesses/>. Zuletzt Besucht: 2016-02-26. 2.4
 - [17] Guarino Ben. The 5 psychological reasons you can’t resist clickbait, 06 2015. URL <https://www.inverse.com/article/4069-the-5-psychological-reasons-you-can-t-resist-clickbait>. Zuletzt Besucht: 2016-01-06. 2.4
 - [18] George Loewenstein. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1):75, 1994. 2.4

- [19] Deana B Davalos, Ruth A Davalos, and Heidi S Layton. Iii. content analysis of magazine headlines changes over three decades? *Feminism & Psychology*, 17(2):250–258, 2007. 2.4
- [20] Barret Grant. *The Official Dictionary of Unofficial English*. McGraw-Hill, 2006. 2.5
- [21] Bram Vijgen. The listicle: An exploring research on an interesting shareable new media phenomenon. *Studia Universitatis Babes-Bolyai-Ephemerides*, (1):103–122, 2014. 2.5
- [22] Maria Konnikova. A list of reasons why our brains love lists. URL <http://www.newyorker.com/tech/elements/a-list-of-reasons-why-our-brains-love-lists>. Zuletzt Besucht: 2016-03-30. 2.5
- [23] Claude Messner and Michaela Wänke. Unconscious information processing reduces information overload and increases product satisfaction. *Journal of Consumer Psychology*, 21(1):9–13, 2011. 2.5
- [24] Merriam Webster. Definition of clickbait. URL <http://www.merriam-webster.com/dictionary/clickbait>. Zuletzt Besucht: 2016-01-11. 3.1
- [25] wikipedia. Clickbait. URL <https://en.wikipedia.org/wiki/Clickbait>. Zuletzt Besucht: 2016-03-30. 3.1
- [26] Jakob Steinschaden. Anatomie des clickbaiting: So ködern uns upworthy, buzzfeed und co. URL <http://www.netzpiloten.de/anatomie-des-clickbaiting-koedern-uns-upworthy-buzzfeed-und-co/>. Zuletzt Besucht: 2016-02-23. 3.1, 3.3
- [27] gruenderszene.de. Call to action. URL <http://www.gruenderszene.de/lexikon/begriffe/call-to-action-cta>. Zuletzt Besucht: 2016-02-23.
- [28] Mindy Lilyquist. Call to action. URL <http://homebusiness.about.com/od/homebusinessglossar1/g/Call-To-Action-Definition.htm>. Zuletzt Besucht: 2016-02-23. 3.1
- [29] Onpage Wiki. Clickbaiting. URL <https://de.onpage.org/wiki/Clickbaiting>. Zuletzt Besucht: 2016-02-23. 3.3
- [30] Hawkings Virgil. Clickbait and stereotypes: Media coverage of the dr congo, nov 2014. URL <http://saccps.blogspot.de/2014/11/clickbait-and-stereotypes-media.html>. Zuletzt Besucht: 2015-10-09. 3.5

- [31] Lucy Benedict. Clickbait, oct 2013. URL <http://putupwithrain.blogspot.de/2013/10/clickbait.html>. Zuletzt Besucht: 2015-10-10.
- [32] mesuckatgames. Misleading title: All gaems are banned!!!!1!!1!, Juli 2015. URL <https://isuckatvideogames.com/2015/07/01/misleading-titles/>. Zuletzt Besucht: 2015-10-10. 3.5
- [33] Dennison Kara. Kill clickbait, 06 2015. URL <http://karadennison.blogspot.de/2015/06/kill-clickbait.html>. Zuletzt Besucht: 2015-10-10. 3.5
- [34] The Chemist. Holy mother of clickbait, feb 2015. URL <http://mychemicaljourney.blogspot.de/2015/02/holy-mother-of-clickbait.html>. Zuletzt Besucht: 2015-10-10. 3.5
- [35] Jon Slattery. Media quotes of the week: From the online chase for clickbait turns journalists into thieves and liars to should star columnists hit sub editors?, apr 2015. URL <http://jonslattery.blogspot.de/2015/04/media-quotes-of-week-from-online-chase.html>. Zuletzt Besucht: 2015-10-10. 3.5
- [36] Steve Hind. My defence of clickbait, nov 2013. URL <http://hindsiteblog.blogspot.de/2013/11/my-defence-of-clickbait.html>. Zuletzt Besucht: 2015-10-10. 3.5
- [37] Surely Sarah. 5 reasons i hate clickbait (and 1 why it works), feb 2014. URL <http://surelysarah.blogspot.de/2014/02/5-reasons-i-hate-clickbait-and-1-why-it.html>. Zuletzt Besucht: 2015-10-10. 3.5
- [38] J Andrew Barta. Let's talk about sex baby. let's talk about a and b., April 2015. URL <https://andrewbarta.wordpress.com/2015/04/20/lets-talk-about-sex-baby-lets-talk-about-a-and-b/>. Zuletzt Besucht: 2015-10-10. 3.5
- [39] Tom Hespos. Why is the internet overrun with clickbait? URL <http://www.the-makegood.com/2015/01/07/why-is-the-internet-overrun-with-clickbait/>. Zuletzt Besucht: 2016-02-23. 3.5
- [40] JESSICAMEHTA. Kim kardashian releases new sex tape! (the problem with clickbait), März 2015. URL <http://mehtafor.com/2015/03/27/kim-kardashian-releases-new-sex-tape-the-problem-with-clickbait/>. Zuletzt Besucht: 2015-10-10. 3.5

- [41] Edward Packard. Six amazing reasons why you should include #historicalclickbait in your teaching, July 2015. URL <https://ucshistory.wordpress.com/2015/07/31/historicalclickbait/>. Zuletzt Besucht: 2015-10-10. 3.5
- [42] twitter.com. Twitter nutzung / fakten zum unternehmen. URL <https://about.twitter.com/de/company>. Zuletzt Besucht: 2016-02-25. 4.1
- [43] Leon Derczynski, Alan Ritter, Sam Clark, and Kalina Bontcheva. Twitter part-of-speech tagging for all: Overcoming sparse and noisy data. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*. Association for Computational Linguistics, 2013. 5.1
- [44] Jeffrey D Ullman, Jure Leskovec, and Anand Rajaraman. Mining of massive datasets, 2011. 5.1
- [45] Edgar Dale and Jeanne S Chall. A formula for predicting readability: Instructions. *Educational research bulletin*, pages 37–54, 1948. 5.1
- [46] Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948. 5.1
- [47] General Inquirer. Welcome to the general inquirer home page. URL <http://www.wjh.harvard.edu/~inquirer>. Zuletzt Besucht: 2015-10-10. 5.1
- [48] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012. 5.2
- [49] Nitesh V Chawla. Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook*, pages 853–867. Springer, 2005. 5.2
- [50] George H John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995. 5.4
- [51] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014. 5.4
- [52] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. 5.4

- [53] Saskia Le Cessie and Johannes C Van Houwelingen. Ridge estimators in logistic regression. *Applied statistics*, pages 191–201, 1992. 5.4
- [54] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006. 5.5, 5.5, 5.5, 5.5
- [55] Eibe Frank and Mark Hall. A simple approach to ordinal classification. 2001. 6.2

Anhang A

Appendix

A.1 Part of speech labels

Die folgende Liste wurde aus erwinkomen.ruhosting.nl/eng/2014_Longdale-Labels.htm Stand: 04.04.2016 entnommen.

ADJP Adjective Phrase. Phrasal category headed by an adjective (including comparative and superlative adjectives). Example: outrageously expensive.

ADVP Adverb Phrase. Phrasal category headed by an adverb (including comparative and superlative adverbs). Examples: rather timidly, very well indeed, rapidly.

CC Coordinating conjunction This category includes 'and', 'but', 'nor', 'or', 'yet' (as in 'Yet it's cheap', 'cheap yet good'), as well as the mathematical operators 'plus', 'minus', 'less', 'times' (in the sense of "multiplied by") and 'over' (in the sense of "divided by"), when they are spelled out. For in the sense of "because" is a coordinating conjunction (CC) rather than a subordinating conjunction (IN). He asked to be transferred, for/CC he was unhappy. So in the sense of "so that", on the other hand, is a subordinating conjunction (IN).

CD Cardinal number

CONJP Conjunction Phrase. Used to mark certain "multi-word" conjunctions, such as 'as well as', 'instead of'.

DT Determiner This category includes the articles a(n), every, no and the, the indefinite determiners another, any and some, each, either (as in either way), neither (as in neither decision), that, these, this and those,

and instances of all and both when they do not precede a determiner or possessive pronoun (as in all roads or both times). (Instances of all or both that do precede a determiner or possessive pronoun are tagged as predeterminers (PDT).) Since any noun phrase can contain at most one determiner, the fact that such can occur together with a determiner (as in the only such case) means that it should be tagged as an adjective (JJ), unless it precedes a determiner, as in such a good time, in which case it is a predeterminer (PDT).

EX Existential there Existential there is the unstressed there that triggers inversion of the inflected verb and the logical subject of a sentence. EX-AMPLES: There/EX was a party in progress.; There/EX ensued a melee.

FRAG Fragment.

FW Foreign word Use your judgment as to what is a foreign word. For me, yoga is an NN, while 'bete noire' and 'persona non grata' should be tagged bete/FW noire/FW and persona/FW non/FW grata/FW, respectively.

IN Preposition or subordinating conjunction We make no explicit distinction between prepositions and subordinating conjunctions. (The distinction is not lost, however - a preposition is an IN that precedes a noun phrase or a prepositional phrase, and a subordinate conjunction is an IN that precedes a clause.) The preposition to has its own special tag TO.

INTJ Interjection. Corresponds approximately to the part-of-speech tag UH.

JJ Adjective Hyphenated compounds that are used as modifiers are tagged as adjectives (JJ). EXAMPLES: happy-go-lucky/JJ; one-of-a-kind/JJ; run-of-the-mill/JJ; Ordinal numbers are tagged as adjectives (JJ), as are compounds of the form n-th X-est, like fourth-largest.

JJR Adjective, comparative Adjectives with the comparative ending -er and a comparative meaning are tagged JJR. More and less when used as adjectives, as in more or less mail, are also tagged as JJR. More and less can also be tagged as JJR when they occur by themselves. Adjectives with a comparative meaning but without the comparative ending -er, like superior, should simply be tagged as JJ. Adjectives with the ending -er but without a strictly comparative meaning ("more X"), like further in further details, should also simply be tagged as JJ.

JJS Adjective, superlative Adjectives with the superlative ending -est (as well as worst) are tagged as JJS. Most and least when used as adjectives, as in the most or the least mail, are also tagged as JJS. Most and least can

also be tagged as JJS when they occur by themselves. Adjectives with a superlative meaning but without the superlative ending -est, like first, last or unsurpassed, should simply be tagged as JJ.

LS List item marker This category includes letters and numerals when they are used to identify items in a list.

LST List marker. Includes surrounding punctuation.

MD – Modal This category includes all verbs that don't take an -s ending in the third person singular present: 'can', 'could', '(dare)', 'may', 'might', 'must', 'ought', 'shall', 'should', 'will', 'would'.

NAC Not a Constituent used to show the scope of certain prenominal modifiers within an NP.

NN Noun, singular or mass

NNS Noun, plural

NNP Proper noun, singular

NNPS Proper noun, plural

NP Noun Phrase. Phrasal category that includes all constituents that depend on a head noun.

NX Used within certain complex noun phrases to mark the head of the noun phrase. Corresponds very roughly to N-bar level but used quite differently.

PDT Predeterminer This category includes the following determiner like elements when they precede an article or possessive pronoun. EXAMPLES: all/PDT his marbles; nary/PDT a soul; both/PDT the girls; quite/PDT a mess; half/PDT his time; rather/PDT a nuisance; many/PDT a moon; such/PDT a good time;

POS Possessive ending The possessive ending on nouns ending in 's or ' is split off by the tagging algorithm and tagged as if it were a separate word. EXAMPLES: John/NNP 's/POS idea; the parents/NNS '/POS distress

PP Prepositional Phrase. Phrasal category headed by a preposition.

PRN Parenthetical.

PRP Personal pronoun This category includes the personal pronouns proper, without regard for case distinctions ('I', 'me', 'you', 'he', 'him', etc.), the reflexive pronouns ending in -self or -selves, and the nominal possessive pronouns 'mine', 'yours', 'his', 'hers', 'ours' and 'theirs'. The adjectival possessive forms 'my', 'your', 'his', 'her', 'its', 'our' and 'their', on the other hand, are tagged PRPS.

PRPS Possessive pronoun This category includes the adjectival possessive forms 'my', 'your', 'his', 'her', 'its', 'one's', 'our' and 'their'. The nominal possessive pronouns 'mine', 'yours', 'his', 'hers', 'ours' and 'theirs' are tagged as personal pronouns (PRP).

PRT Particle. Category for words that should be tagged RP.

QP Quantifier Phrase (i.e. complex measure/amount phrase); used within NP.

RB Adverb This category includes most words that end in -ly as well as degree words like quite, too and very, posthead modifiers like enough and indeed (as in good enough, very well indeed), and negative markers like 'not', 'n't' and 'never'.

RBR Adverb, comparative Adverbs with the comparative ending -er but without a strictly comparative meaning, like later in We can always come by later, should simply be tagged as RB.

RBS Adverb, superlative

RP Particle This category includes a number of mostly monosyllabic words that also double as directional adverbs and prepositions.

RRC Reduced Relative Clause.

S Simple declarative clause. One that is not introduced by a (possible empty) subordinating conjunction or a wh-word and that does not exhibit subject-verb inversion.

SBAR Clause introduced by a (possibly empty) subordinating conjunction. Direct question introduced by a wh-word or a wh-phrase. Indirect questions and relative clauses should be bracketed as SBAR, not SBARQ.

SINV Inverted declarative sentence One in which the subject follows the tensed verb or modal.

SQ Inverted yes/no question, or main clause of a wh-question, following the wh-phrase in SBARQ.

- SYM** Symbol This tag should be used for mathematical, scientific and technical symbols or expressions that aren't words of English. It should not be used for any and all technical expressions. For instance, the names of chemicals, units of measurements (including abbreviations thereof) and the like should be tagged as nouns.
- TO** to To is tagged TO, regardless of whether it is a preposition or an infinitival marker
- UCP** Unlike Coordinated Phrase
- UH** Interjection This category includes 'my' (as in 'My, what a gorgeous day'), 'oh', 'please', 'see' (as in 'See, it's like this'), 'uh', 'well' and 'yes', among others.
- VB** Verb, base form This tag subsumes imperatives, infinitives and subjunctives. EXAMPLES: Imperative: Do/VB it. EXAMPLES: Infinitive: You should do/VB it.; We want them to do/VB it.; We made them do/VB it.; EXAMPLES: Subjunctive: We suggested that he do/VB it.
- VBD** Verb, past tense This category includes the conditional form of the verb to be. EXAMPLES: If I were/VBD rich, ... ; If I were/VBD to win the lottery, ...
- VBG** Verb, gerund or present participle
- VCN** Verb, past participle
- VBP** Verb, non-3rd person singular present
- VBZ** Verb, 3rd person singular present
- VP** Verb Phrase. Phrasal category headed a verb.
- WDT** Wh-determiner This category includes which, as well as that when it is used as a relative pronoun.
- WHADJP** Wh-adjective Phrase. Adjectival phrase containing a wh-adverb, as in 'how hot'.
- WHADV** Wh-adverb Phrase. Introduces a clause with an NP gap. May be null (containing the 0 complementizer) or lexical, containing a wh-adverb such as 'how' or 'why'.

WHNP Wh-noun Phrase. Introduces a clause with an NP gap. May be null (containing the 0 complementizer) or lexical, containing some wh-word, e.g. 'who', 'which book', 'whose daughter', 'none of which', or 'how many leopards'.

WHPP Wh-prepositional Phrase. Prepositional phrase containing a wh-noun phrase (such as 'of which' or 'by whose authority') that either introduces a PP gap or is contained by a WHNP.

WP Wh-pronoun This category includes 'what', 'who' and 'whom'.

WPS Possessive wh-pronoun This category includes the wh-word 'whose'

WRB Wh-adverb This category includes 'how', 'where', 'why', etc. When in a temporal sense is tagged WRB. In the sense of if", on the other hand, it is a subordinating conjunction (IN). EXAMPLES: When/WRB he finally arrived, I was on my way out.; I like it when/IN you make dinner for me.

X Unknown, uncertain, or unbracketable. X is often used for bracketing typos and in bracketing 'the...the'-constructions.

. ; ? ! - Punctuation mark, sentence close: ' ' , , , ' ? ' , !

, - Punctuation mark, comma: ', '.

: - Punctuation mark, colon: ':'.

LRB Contextual separator, left parenthesis

RRB Contextual separator, right parenthesis

SQT Start quote

EQT End quote

A.2 χ^2 -Verfahren Top 100 Attribute

Ersteller	Link 2 Zeichen: ne	Link 3 Zeichen: the
Tweet POST 1 Worte:	I;You	Link 3 Zeichen: you
NNP	Link 1 Zeichen: C	Link 2 Zeichen: ex
Tweet Einfache Wörter	Link 1 Zeichen: f	Link 2 Zeichen: om
Link 1 Zeichen: l	Tweet 2 Zeichen: yo	Link 3 Zeichen: ate
Tweet Lesbarkeit	Tweet Pronomen	Link 1 Zeichen: T
Tweet Zeichenlänge	Link 1 Worte: reported	Link 3 Zeichen: eve
Link 1 Zeichen: s	Link 3 Zeichen: a	Tweet POST 2 Worte:
Tweet Stopworte	Link 1 Worte: a	IN NNP
Tweet 1 Zeichen: ?	Link 2 Zeichen: a	Tweet POST 1 Worte:
Link 1 Worte: in	T;TargetCharacterLength	HT
Link 3 Zeichen: in	Link 2 Zeichen: en	Link 2 Zeichen: es
Link 3 Zeichen: epo	Link 2 Zeichen: th	Link 3 Zeichen: Tue
Link 1 Zeichen: n	Link 1 Zeichen: r	Link 2 Zeichen: la
Link 1 Zeichen: t	Link 2 Zeichen: ar	Link 3 Zeichen: aid
Link 1 Zeichen: x	Tweet 1 Zeichen: a	Link 3 Zeichen: nte
Link 1 Zeichen: u	Link 1 Zeichen: d	Tweet POST 2 Worte: :
Link Teilungen	Link 2 Zeichen: ow	NNP
Link 1 Worte: is	Link 2 Zeichen: of	Tweet POST 1 Worte:
Link 3 Zeichen: is	Link 2 Zeichen: ev	WRB
Link 2 Zeichen: ee	Link 3 Zeichen: esd	Tweet Präposition
Link 3 Zeichen: sai	Link 3 Zeichen: Off	Link 1 Zeichen: .
Link 2 Zeichen: ur	Link 1 Zeichen: :	Tweet 1 Zeichen: #
Link 2 Zeichen: an	Link 2 Zeichen: ba	Link 1 Zeichen: D
Link 1 Worte: police	Link 2 Zeichen: se	Link 2 Zeichen: ha
Link 2 Zeichen: ś	Link 3 Zeichen: Jun	Link 2 Zeichen: ou
Link 2 Zeichen: mo	Tweet POST 1 Worte:	Link 1 Zeichen: ,
Tweet 3 Zeichen: you	DT	Tweet POST 1 Worte: :
Link 2 Zeichen: na	Link 3 Zeichen: tat	Tweet 1 Worte: you
Link 1 Zeichen: o	Link 2 Zeichen: sd	Tweet 1 Zeichen: r
Link 2 Zeichen: at	Link 2 Zeichen: wi	Link 3 Zeichen: rep
Tweet Mittlere Wortlän-	Link 2 Zeichen: te	Link 1 Zeichen: y
ge	Link 2 Zeichen: y,	Link 1 Zeichen: c
Link 2 Zeichen: le	Link Lesbarkeit	
Link 1 Worte: said	Link 1 Worte: State	
Link 3 Zeichen: sda	Link 2 Zeichen: in	
Link 1 Zeichen: e	Link 3 Zeichen: rte	

A.3 Selbsterstellte Attributmenge

Tweet Beginnt mit	Tweet 1 Zeichen: n	Link 1 Worte: said
Zeichen	Tweet 1 Zeichen: r	Link 1 Worte: Share
Tweet Einfache Worte	Tweet 1 Zeichen: S	Link 1 Worte: This
Tweet lesbarkeit	Tweet 2 Worte: in the	Link 1 Worte: What
Tweet POS-Tag: : NNP	Tweet 2 Worte: is the	Link 1 Worte: you
Tweet POS-Tag: DT	Tweet 2 Worte: of the	Link 1 Worte: You
Tweet POS-Tag: IN	Tweet 2 Worte: on the	Link 1 Worte: your
Tweet POS-Tag: IN	Tweet 2 Worte: to be	Link 2 Worte: contribu-
NNP	Tweet 2 Worte:	ted to
Tweet POS-Tag: NN IN	WATCH:	Link 2 Worte: got a
Tweet POS-Tag: NNP	Tweet 2 Worte: ...	Link 2 Worte: how to
Tweet POS-Tag: NNP	Tweet 2 Zeichen: @t	Link 2 Worte: if you
NNP	Tweet 2 Zeichen: am	Link 2 Worte: If you
Tweet POS-Tag: NNP	Tweet 2 Zeichen: er	Link 2 Worte: of your
VBZ	Tweet 2 Zeichen: hy	Link 2 Worte: with your
Tweet POS-Tag: PRP	Tweet 2 Zeichen: in	Link 2 Worte: you can
Tweet POS-Tag: WRB	Tweet 2 Zeichen: la	Link 2 Worte: you should
Tweet Präposition	Tweet 2 Zeichen: rs	Link 2 Worte: your own
Tweet Zeichenlänge	Tweet 2 Zeichen: Th	Link 3 Worte: 2015 Face-
Tweet 1 Worte: as	Tweet 2 Zeichen: th	book 0
Tweet 1 Worte: How	Tweet 2 Zeichen: yo	Link 3 Worte: contribu-
Tweet 1 Worte: in	Tweet 3 Zeichen: @th	ted to this
Tweet 1 Worte: most	Tweet 3 Zeichen: hap	Link 3 Worte: emailed
Tweet 1 Worte: on	Tweet 3 Zeichen: his	the post.
Tweet 1 Worte: This	Tweet 3 Zeichen: How	Link 3 Worte: Facebook
Tweet 1 Worte: this	Tweet 3 Zeichen: in	0 Twitter
Tweet 1 Worte: Why	Tweet 3 Zeichen: mos	Link 3 Worte: have suc-
Tweet 1 Worte: you	Tweet 3 Zeichen: Thi	cessfully emailed
Tweet 1 Worte: your	Tweet 3 Zeichen: thi	Link 3 Worte: success-
Tweet 1 Zeichen: ?	Tweet 3 Zeichen: Why	fully emailed the
Tweet 1 Zeichen: a	Tweet 3 Zeichen: you	Link 3 Worte: The Huf-
Tweet 1 Zeichen: C	Link 1 Worte: 4.	fington Post
Tweet 1 Zeichen: E	Link 1 Worte: can	Link 3 Worte: You have
Tweet 1 Zeichen: i	Link 1 Worte: Jun	successfully
Tweet 1 Zeichen: l	Link 1 Worte: of	
Tweet 1 Zeichen: M	Link 1 Worte: or	