

Hochschule Rhein-Waal
Fakultät Kommunikation und Umwelt
Prof. Dr.-Ing. Ulrich Greveler
Prof. Dr. Thomas Richter

Untersuchung der Vorgehensweise ausgewählter Internetplattformen
gegen gefälschte Online-Bewertungen

Bachelorarbeit
im Studiengang
Medien- und Kommunikationsinformatik
zur Erlangung des akademischen Grades

Bachelor of Science

vorgelegt von
Johannes Karl Nolte
johannes-karl.nolte [at] hsrw.org

Abgabedatum: 20. Februar 2018

Zusammenfassung

Diese Arbeit dokumentiert ein Experiment zur Untersuchung des Umgangs von Facebook, Google Maps, Google Play Store und Zalando mit gefälschten Bewertungen („Opinion Spam“). Es wurde untersucht, ob die Plattformen im Rahmen dieser Arbeit abgegebene gefälschte Bewertungen veröffentlichen. Dazu wurden je Plattform 24 unterschiedlich authentisch wirkende Konten erstellt. Mit diesen wurden je 48 gefälschte Bewertungen nach folgenden Kriterien abgegeben: 1) Kopierte Produkt-/Ortsbeschreibung oder Werbesprache 2) Aussagen ohne Produkt-/Ortsbezug oder inhaltliche Fehler 3) Blindtexte oder sehr viele Rechtschreibfehler 4) eine Meinung verknüpft mit der Aussage, dass der Bewertende das Produkt/den Ort nicht kennt. Innerhalb von 14 Tagen wurde mehrmals überprüft, ob eine Bewertung freigeschaltet oder gelöscht wurde.

Bei Facebook wurden 6,7% der abgegebenen Bewertungen innerhalb der ersten 14 Tage gelöscht. Nachdem die verbleibenden Bewertungen der Plattform gemeldet wurden, waren es 47,7%. Bei Google Maps wurden 4,2% entfernt, im Google Play Store 12,5% (bei beiden keine Änderung durch Melden). Bei Zalando (kein Melden möglich) war die eigenständige Löschrquote mit 43,8% mit Abstand am höchsten. Besonders häufig wurden im Experiment Blindtexte gelöscht. Positive und negative Bewertungen wurden nahezu gleich oft entfernt. Bewertungen von Konten mit simulierter Nutzerinteraktion wurden häufiger gelöscht als Bewertungen von Konten ohne Nutzerinteraktion.

Das Experiment hat gezeigt, dass sehr viele gefälschte Bewertungen, die eindeutig den Richtlinien der Plattformen widersprechen, nicht gelöscht wurden. Deren Sichtbarkeit variiert allerdings je nach Plattform und Produkt/Ort stark. Der Aufwand für die Erstellung vieler Konten bei den Diensten ist überschaubar. Wenn Nutzer gefälschte Bewertungen nach den im Experiment untersuchten Kriterien abgeben wollen, werden sie von den Plattformen daran wenig gehindert.

Schlüsselwörter:

Gefälschte Bewertungen, Opinion Spam, Google, Facebook, Zalando

Abstract

This bachelor thesis contains the documentation of an experiment to evaluate how Facebook, Google Maps, Google Play Store and Zalando deal with fake reviews (“opinion spam”). Measured was whether the platforms publish or delete fake reviews submitted in the context of this work. For this purpose, 24 accounts with different attributes according to age and user interaction were created on each platform. With those accounts 48 reviews per platform were submitted based on the following criteria: 1) Copied product/location description or advertising-language 2) reviews without product/location reference or contextual errors 3) Blind texts or many spelling mistakes 4) reviews which contain an opinion but state that the author does not know the product/location.

6.7% of the reviews submitted to Facebook were deleted within the first 14 days, 47.7% after reporting the remaining reviews to Facebook. Google Maps deleted 4.2%, Google Play Store 12.5% (both showed no changes after reporting). Zalando (no reporting possible) has the by far highest independent deletion rate with 43.8%. Blind texts were most often deleted during the experiment. Positive and negative ratings were deleted almost the same number of times. Reviews from accounts with simulated user interaction were deleted more often than reviews from accounts without user interaction.

The experiment showed that many fake reviews, which contradict the guidelines of the platforms, were not deleted. But their visibility varies depending on platform and product/location. The creation of many accounts at the services does not require too much effort. If users want to publish fake reviews according to the criteria used in the experiment, the platforms do not really hinder them.

Key words:

Fake Reviews, Opinion Spam, Google, Facebook, Zalando

Inhaltsverzeichnis

1	Einleitung und Motivation	1
2	Hintergrund.....	3
2.1	„Gefälschte“ Bewertungen	3
2.2	Ausmaß gefälschter Bewertungen	4
2.3	Erkennung gefälschter Bewertungen	5
2.4	Untersuchungen zum Verfassen gefälschter Bewertungen.....	7
2.5	Vorstellung der Plattformen.....	8
2.5.1	Facebook	8
2.5.2	Google Maps.....	9
2.5.3	Google Play Store.....	9
2.5.4	Zalando.....	10
2.6	Regeln der Plattformen gegen gefälschte Bewertungen.....	10
2.6.1	„Gemeinschaftsstandards“ bei Facebook	11
2.6.2	„Richtlinien für Rezensionen“ bei Google Maps.....	11
2.6.3	„Richtlinien zum Posten von Kommentaren“ im Google Play Store	11
3	Methodik und Vorgehensweise	13
3.1	Auswahl der Plattformen	13
3.2	Kontrolle der Folgen der abgegebenen Bewertungen.....	13
3.3	Klassifikation der Konten und Bewertungen.....	14
3.3.1	Konten	14
3.3.2	Bewertungen	15
3.4	Erstellung der Konten	18
3.5	Browserumgebungen.....	19
3.6	Erstellung der Bewertungen.....	20
3.7	Kontrolle des Online-Status.....	21
3.8	Reflexion der Methodik	21
3.8.1	Störvariablen.....	22
3.8.2	Validität	23

3.8.3	Fazit zur Reflexion der Methodik.....	23
4	Auswertung.....	24
4.1	Auswertung des Experiments.....	24
4.1.1	Facebook	24
4.1.2	Google Maps.....	25
4.1.3	Google Play Store.....	26
4.1.4	Zalando.....	26
4.1.5	Plattformübergreifende Auswertung.....	28
4.1.6	Sichtbarkeit der abgegebenen Bewertungen.....	31
4.2	Folgeexperiment: Melden der Bewertungen.....	32
4.3	Googles Registrierungsprozess.....	35
5	Zusammenfassung.....	36
5.1	Erstellung der Konten	36
5.2	Verfassen von Bewertungen	36
5.3	Methodik.....	38
6	Fazit	39
7	Ausblick.....	41
8	Literaturverzeichnis.....	43
	Anhang	46

Abbildungsverzeichnis

Abbildung 1: Anzahl und Anteil der ungelöschten und gelöschten Bewertungen bei Zalando ..	28
Abbildung 2: Anzahl der plattformübergreifend gelöschten und ungelöschten Bewertungen ..	29
Abbildung 3: Löschrquote der abgegebenen Bewertungen nach Plattformen	29
Abbildung 4: Plattformübergreifende Löschung aufgeschlüsselt nach Klassen der Konten	30
Abbildung 5: Ein Blindtext als Bewertung eines Kinderfilms im Google Play Store wurde zehnmal als hilfreich bewertet	32
Abbildung 6: Einen Richtlinienverstoß bei Google Maps melden.....	33
Abbildung 7: Facebook ermahnt den Autor einer gefälschten Bewertung.....	34

Tabellenverzeichnis

Tabelle 1: Klassen der Konten.....	15
Tabelle 2: Klassen der Bewertungen.....	18

Hinweis: Zur Verbesserung der Lesbarkeit wurde in der vorliegenden Arbeit auf eine geschlechtsspezifische Differenzierung verzichtet. Die verwendeten Bezeichnungen (bspw. „der Nutzer“) sind nicht als geschlechtsspezifisch zu verstehen.

1 Einleitung und Motivation

Eine der wesentlichen Eigenschaften des Internets ist es, dass dessen Nutzer nicht nur passiv Inhalte vielfältiger Art konsumieren können, sondern auch aktiv selbst Inhalte erstellen können. So sind unzählige Communities und Plattformen entstanden, auf denen sich die Nutzer gegenseitig informieren und unterhalten. Eine Form der gegenseitigen Information sind online verfasste Bewertungen. Diese können etwa Meinungen zu Produkten, Dienstleistungen oder Orten enthalten. Online-Bewertungen sind abzugrenzen von redaktionellen Tests. Diese sind in der Regel wesentlich umfangreicher als Online-Bewertungen. Letztere stammen von Benutzern, deren Hauptrolle beispielsweise die eines Käufers ist und die die Bewertung zusätzlich verfassen.

Besonders verbreitet sind Bewertungen von Restaurants und Hotels (etwa auf den Plattformen Yelp, TripAdvisor und booking.com) sowie von Produkten (beispielsweise Amazon und andere Onlineshops). Auch Ärzte, Hochschulen, Arbeitgeber und diverse andere Branchen, Orte und Organisationen werden in den letzten Jahren immer öfter online bewertet (Ott, 2011, S. 1). So ist beispielsweise die Anzahl der Bewertungen auf Yelp zwischen 2009 und 2016 von rund 8,8 Millionen auf über 121 Millionen angestiegen (Statista, 2016a).

Grundsätzlich kann zwischen zwei Formen von Bewertungsportalen unterschieden werden. Die reinen Bewertungsportale bieten ausschließlich und oft werbefinanziert Nutzern die Möglichkeit, Bewertungen für eine Branche bzw. Waren oder Dienstleistungen zu verfassen. Oft gibt es einen thematischen Schwerpunkt, beispielsweise Ärzte (etwa bei Jameda) oder Gastronomie (etwa bei Yelp). Auf der anderen Seite stehen Webseiten wie beispielsweise Online-Shops und Buchungsplattformen, die Bewertungen einsetzen, um ihren Kunden die Kaufentscheidung zu erleichtern oder eine andere Dienstleistung aufzuwerten. In diese Kategorie fallen etwa Amazon, Expedia und Google Maps. Von Nutzern verfasste Bewertungen haben dabei für Verbraucher einen hohen Stellenwert. Über 80% der 2016 im Auftrag der HolidayCheck Group befragten Internetnutzer gaben an, dass Bewertungen Einfluss auf ihre Kaufentscheidung haben (Statista, 2016b). Ähnlich viele Nutzer sind sich jedoch auch bewusst, dass Online-Bewertungen gefälscht sein können, zeigt eine weitere Studie bezogen auf Hotel-Bewertungen (Mühlenbeck und Skibicki, 2010, S. 62). So ist die paradoxe Situation entstanden, dass Internetnutzer sich stark nach Bewertungen richten, obwohl sehr vielen die Möglichkeit der Fälschung dieser bekannt ist. Wenn die Bewertungsplattformen es nicht schaffen, gefälschte Bewertungen zu löschen, ist es in vielen Fällen wahrscheinlich, dass in der Folge Nutzer getäuscht werden.

Um die Anfälligkeit von vier ausgewählten Plattformen für gefälschte Online-Bewertungen zu überprüfen, untersucht das Experiment im Rahmen der Arbeit den Online-Status (freigeschaltet/gelöscht) von gefälschten Bewertungen. Dazu wurde bei vier Plattformen eine größere Anzahl von gefälschten Bewertungen nach festgelegten Kriterien abgegeben. Auch die Erstellung von Konten nach verschiedenen Kriterien bei den untersuchten Plattformen gehörte zum Versuchsablauf. Dies geschah unter der Annahme, dass die Plattformen mit Bewertungen von Benutzern, deren Konten schon länger aktiv sind, anders umgehen als mit Bewertungen, hinter denen etwa ganz neue Konten stehen. Bei der Abgabe der gefälschten Bewertungen wurden verschiedene reale Motivationen für gefälschten Bewertungen imitiert. Beispielsweise wurden Bewertungen mit einem Text ohne Bezug zum bewerteten Produkt oder Ort (Bewertungsgegenstand) abgegeben um vermeintlich schnell die durchschnittliche Bewertung zu beeinflussen. Auch reine Spam-Bewertungen mit Blindtext wurden abgegeben um zu testen, ob es den Plattformen gelingt, nicht hilfreiche Bewertungen herauszufiltern.

Durch den beschriebenen Versuchsablauf konnte untersucht werden, wie die Plattformen mit gefälschten Bewertungen umgehen und wie gut sie der Verantwortung gegenüber ihren Nutzern nachkommen, möglichst authentische Bewertungen anzuzeigen. Dies wurde nach Kenntnis des Autors in diesem Umfang sowie bei den ausgewählten Plattformen bisher nicht untersucht. Es handelt sich jedoch um einen entscheidenden Faktor bei der Beurteilung der Qualität von Bewertungen, auf die sehr viele Menschen täglich zugreifen.

2 Hintergrund

Bevor die Durchführung und die Ergebnisse des Experiments dargestellt werden, werden im Folgenden zunächst Grundlagen zu gefälschten Bewertungen sowie zu den untersuchten Plattformen erläutert.

2.1 „Gefälschte“ Bewertungen

Gefälschte Bewertungen sind nicht immer eindeutig von "ungefälschten" zu unterscheiden. So gibt es eine Reihe von Grenzfällen, beispielsweise wenn der Bewertende in einer engeren Beziehung zum Anbieter des Bewertungsgegenstandes steht. Möglicherweise kennt er zwar den Bewertungsgegenstand, seine Objektivität bei dessen Bewertung kann aber angezweifelt werden. Eindeutig sind Fälle, in denen der Bewertende gar nicht mit dem Bewertungsgegenstand in Kontakt kommt, etwa weil es sich um über eine Agentur gebuchte und bezahlte Bewertungen handelt. Schwierig einzuordnen sind vermeintlich unabhängige aber mit Anreizen wie kostenlosen Produkten initiierte Bewertungen, etwa im „Amazon Vine“-Programm. Dort ist die Objektivität oft nicht gegeben (Kuhr, 2013). Als nicht gefälscht betrachtet werden können Bewertungen, denen ein echtes Erfahrungserlebnis zugrunde liegt, die jedoch zu einem objektiv betrachtet ungerechtfertigten Fazit kommen. Es handelt sich trotzdem um individuelle Eindrücke/Meinungen. Die Frage, ob eine Bewertung gefälscht ist, lässt sich also oft viel mehr in ihrem Entstehungshintergrund als mit ihrem Wortlaut beantworten.

Gefälschte Bewertungen können sowohl positiv, wie auch negativ ausfallen. Während positive gefälschte Bewertungen der Verkaufsförderung dienen und in der Regel durch eine Person initiiert werden, die von höheren Absätzen oder höherer Reputation profitiert (etwa Händler, Inhaber oder Autor), dienen negative gefälschte Bewertungen dazu, Konkurrenten zu schaden und sich so indirekt Vorteile zu verschaffen.

Gefälschte Bewertungen werden in der englischsprachigen Literatur häufig als „opinion spam“ bezeichnet. Bing Liu definiert diesen Begriff wie folgt:

Opinion spam refers to fake or bogus opinions that try to deliberately mislead readers or automated systems by giving undeserving positive opinions to some target objects in order to promote the objects and/or by giving malicious negative opinions to some other objects in order to damage their reputations. (Liu, 2010, S. 2)

In dieser Arbeit bezieht sich der Begriff „gefälschte Bewertungen“ auf bewusst nicht wahrheitsgemäß verfasste Bewertungen, die keine persönliche Meinung ihres Autors zum Bewertungsgegenstand beinhalten und damit den Zweck einer Nutzerbewertung verfehlen. Im

Experiment wurden beispielsweise verschiedene Arten von Texten ohne Produktbezug abgegeben. Näheres dazu ist in Kapitel 3.3 erläutert.

Wie bereits in der Einleitung dargelegt, erhöhen positive gefälschte Bewertungen die Kaufbereitschaft für ein Produkt oder eine Dienstleistung. Da dies jedoch Käufe, Buchungen, Besuche etc. durch potenziell kritische Bewertungsautoren zur Folge hat, ist damit zu rechnen, dass in der Folge unabhängige Bewertungen verfasst werden. So ist es in der Regel nicht möglich, einen schlechten Bewertungsgegenstand langfristig nur mit gefälschten positiven Rezensionen erfolgreich darzustellen. Die wachsende Anzahl negativer Bewertungen wird die gefälschten positiven Bewertungen beim Erreichen einer kritischen Masse unglaubwürdig erscheinen lassen.

2.2 Ausmaß gefälschter Bewertungen

Wie im vorherigen Kapitel 2.1 erläutert, ist die Entscheidung, wann eine Bewertung als gefälscht klassifiziert werden kann, nicht trivial. Dazu kommt, dass im Normalfall nur der Autor und ggf. Auftraggeber einer Bewertung sicher weiß, ob die Bewertung echt oder gefälscht ist. Die Plattformen sowie Dritte können dies nur anhand verschiedener Kriterien mutmaßen und je nach Eindeutigkeit der eigenen Einschätzung entscheiden. Dies macht auch die Forschung zu diesem Thema schwierig: Etwa zum Erstellen von Trainingsdaten für Erkennungs-Algorithmen müssen echte Bewertungen manuell klassifiziert werden, wobei jedoch Fehlentscheidungen getroffen werden können. Damit Forscher sicherstellen können, dass sie wirklich mit gefälschten Bewertungen arbeiten, müssen sie diese selbst schreiben (lassen). Dann handelt es sich jedoch nicht mehr um echte gefälschte Bewertungen, sondern um möglicherweise nicht authentisch gefälschte Bewertungen. Mukherjee et al. (2013) haben gezeigt, dass sich extra für ein Experiment gefälschte Bewertungen deutlich leichter erkennen lassen als echte, von Internetnutzern gefälschte Bewertungen, die in diesem Fall von Yelps Filter-Algorithmus klassifiziert wurden. Untersuchungen, die messen wollen, wie viele Online-Bewertungen gefälscht sind, konnten dies somit nicht direkt messen. Stattdessen basieren sie in der Regel auf vielen einzelnen Vermutungen, ob Bewertungen gefälscht sind.

In einer Studie der Fachhochschule Worms und der itb Berlin aus dem Jahr 2014 wurden 15 Hotelbewertungsplattformen nach ihrer Einschätzung zur Fälschungsquote von Bewertungen auf der eigenen Seite befragt. Sieben Plattformen gaben Werte unter einem Prozent an, sieben weitere zwischen 1% und 5%. Lediglich Yelp schätzte die Quote von gefälschten Bewertungen auf der eigenen Seite auf 25% (Conrady, 2014, S. 38). Möglicherweise wurden hier als gefälscht klassifizierte Bewertungen mitgezählt, die Yelp jedoch mit dem zusätzlichen Hinweis „Beiträge, die momentan nicht empfohlen werden“ unter den vollständig freigeschalteten Bewertungen anzeigt.

Mukherjee (2015, S. 21) schätzt, dass bis zu 30% der Bewertungen im Internet gefälscht sein könnten. Fundiertere und deutlich niedrigere Zahlen liefern Ott, Cardie und Hancock (2012) bezogen auf sechs Hotelbewertungsportale. Mit ihrem Versuchsaufbau attestieren sie den untersuchten Plattformen jeweils ähnlich hohe Fälschungsquoten von rund 2%. Lediglich bei TripAdvisor stieg die Zahl der als gefälscht klassifizierten Bewertungen im Zeitraum der Untersuchung auf rund 6% an. Allgemein halten die Autoren fest, dass der Anteil der gefälschten Bewertungen mit der Offenheit einer Plattform zusammenhängt. Je höher die Hürden dafür sind eine Bewertung abzugeben (z. B. Buchung einer Übernachtung vor der Bewertung), desto niedriger ist der Anteil gefälschter Bewertungen.

2.3 Erkennung gefälschter Bewertungen

Facebook, Google und Zalando haben keine Informationen dazu veröffentlicht, wie sie Bewertungen untersuchen, um deren Authentizität zu bewerten. Anders sieht es auf der Webseite des Dienstes ReviewMeta (www.reviewmeta.com) aus, wo ausführliche Informationen zur Funktionsweise der Erkennung gefälschter Bewertungen zu finden sind. ReviewMeta ist ein Online-Tool, das Bewertungen auf Amazon US/CA/UK/DE/FR/IT/AU sowie Bodybuilding.com analysiert und den Nutzer über die vermutete Glaubwürdigkeit der Bewertungen eines Produktes informiert. Dies geschieht, indem eine angepasste Amazon-Bewertungsübersicht (Sterne, Anzahl Bewertungen, Kategorie-Bewertungsrang) angezeigt wird, welche nur noch aus Rezensionen besteht, die ReviewMeta für glaubwürdig hält. Der Dienst arbeitet dabei mit 15 Faktoren, welche vom Nutzer auch individuell gewichtet werden können. In den Default-Einstellungen zählt beispielsweise eine Bewertung, deren Autor hauptsächlich Produkte einer einzigen Marke bewertet hat, als 100% unglaubwürdig und eine Bewertung ohne verifizierten Kauf als 40% unglaubwürdig. Weitere Kriterien von ReviewMeta sind: Durchschnittsnote und Anzahl der Rezensionen eines Autors, Zeitstempel der Bewertungen, initiierte Bewertungen (z. B. Amazon Vine Programm), Überschneidungen in der Liste der von einem Autor bewerteten Produkte mit der anderer Autoren, durchschnittliche Wortanzahl einer Bewertung verglichen mit dem Durchschnitt der Produktkategorie sowie der Anzahl der vom Plattformanbieter gelöschten Bewertungen eines Autors (ReviewMeta, 2017). Diese Kriterien lassen sich nicht vollständig auf andere Plattformen übertragen, es ist aber plausibel, dass andere Plattformen ähnliche Kriterien zur Untersuchung von Bewertungen nutzen.

Noch detailliertere Informationen darüber, wie gefälschte Bewertungen erkannt werden können, sind in wissenschaftlichen Publikationen zu dieser Fragestellung zu finden. Verschiedene Wissenschaftler haben in den vergangenen Jahren an unterschiedlichen Methoden geforscht, um gefälschte Bewertungen zu erkennen.

Mukherjee, Liu und Glance (2012) haben sich zur Erkennung von gefälschten Bewertungen auf die Erkennung von Gruppen von Bewertungsfälschern („spammer groups“) konzentriert. Die Erkennung von solchen Gruppen am Beispiel von Amazon-Bewertungen erfolgt dabei zunächst durch eine Analyse auf Überschneidungen in der Liste der bewerteten Produkte der Autoren von Bewertungen. Anschließend wurden acht Faktoren überprüft und daraus berechnet, wie hoch die Wahrscheinlichkeit ist, dass es sich um eine Bewertung eines Mitglieds einer Spammer Gruppe handelt. Dazu zählen beispielsweise Korrelationen beim Zeitpunkt der Bewertungsabgabe und Ähnlichkeit bei der Anzahl der vergebenen Sterne sowie der verwendeten Wörter. Zusätzlich wurden die Bewertungen individuell untersucht, unter anderem ob eine Bewertung als eine der ersten für ein Produkt veröffentlicht wurde und ob es innerhalb der Bewertungen eines Autors Ähnlichkeiten in der Anzahl der vergebenen Sterne und der Wortwahl gibt. Die Erkennung von gefälschten Bewertungen anhand dieser kombinierten, teils gruppenspezifischen Faktoren wird als erfolgreicher beschrieben, als die Erkennung von einzelnen gefälschten Bewertungen. Da eine Autorin des Papers Angestellte von Google ist, ist es möglich, dass Erkenntnisse aus der Forschung bei Google aufgegriffen wurden.

In der Arbeit von Lim et al. (2010) wurden gefälschte Bewertungen dadurch erkannt, dass ihre Meinung von anderen Rezensionen abweicht und sich der Autor einer Bewertungen auf eine bestimmte Produktgruppe konzentriert. Der Einfluss von gefälschten Bewertungen ist demnach höher auf das Durchschnittsrating als der von Rezensionen, welche von Nutzern als „Nicht hilfreich“ bewertet wurden. Einen ähnlichen Ansatz verfolgen Wu et al. (2010), die den Faktor des Verzerrens („disortion“) zum Erkennen von gefälschten Bewertungen auf TripAdvisor vorstellen. Dabei wird verglichen, wie sich die Löschung einer Bewertung auf die durchschnittliche Bewertung eines Ortes auswirkt, verglichen mit der Löschung einer zufälligen Bewertung. Dies kann helfen, die Glaubwürdigkeit einer Bewertung zu bestimmen.

Die vorgestellten Quellen zeigen, dass die Plattformen durchaus eine Chance haben, gefälschte Bewertungen zu erkennen. Die meisten Autoren beschreiben ihre entwickelten Erkennungsmethoden als erfolgreich. Grundsätzlich ist aber anzumerken, dass sich die Forschungsarbeiten in der Regel nur auf eine Plattform oder einen Typ von Online-Bewertungen (z. B. Hotel-Bewertungen) beziehen. Die Erkenntnisse der Arbeiten können eventuell nicht auf andere Arten von Bewertungen übertragbar sein oder sich im produktiven Einsatz mit einer Vielzahl von nutzergenerierten Bewertungen anders als im Forschungskontext verhalten.

2.4 Untersuchungen zum Verfassen gefälschter Bewertungen

Wissenschaftliche Untersuchungen zum Thema „Gefälschte Bewertungen“ beziehen sich vor allem auf das Thema der algorithmischen Erkennung gefälschter Bewertungen. So sind viele Erkenntnisse aus Sicht der Plattformen, die eingehende Bewertungen auf ihre Authentizität prüfen wollen, vorhanden. Dagegen ist die Zahl der Publikationen, die sich mit dem Thema beschäftigen, indem sie sich mit der Herangehensweise der Autoren von gefälschten Bewertungen auseinandersetzen, sehr gering. Da in dieser Arbeit gefälschte Bewertungen abgegeben wurden, werden im Folgenden drei Veröffentlichungen mit dieser Perspektive vorgestellt.

Theodoros Lappas (2012) stellt in einem Konferenzpapier die „böartige“ Perspektive („malicious perspective“) auf gefälschte Bewertungen vor. Dazu wurden Kriterien entwickelt, die für den Erfolg einer gefälschten Bewertung entscheidend sind. Dabei handelt es sich um die Authentizität, erreicht etwa durch geringe Abweichung zu anderen Bewertungen und eine gute Lesbarkeit, sowie um den Einfluss einer Bewertung. Letzterer wird nicht nur anhand der Sterne-Bewertung festgestellt, sondern basierend auf verschiedenen innerhalb des Freitextes bewerteten Produkteigenschaften. Lappas vergleicht die positive Bewertung des Zielproduktes und die negative Bewertung von Konkurrenzprodukten als Angriffsmethoden auf einen Korpus von Online-Bewertungen. Dazu wurde ein Datensatz aus echten Amazon-Bewertungen verwendet und berechnet, wie aufwendig es wäre, das Meinungsbild zu einem Produkt zu verzerren. Ergebnis ist, dass das Verfassen negativer Bewertungen für Konkurrenzprodukte bei den Produkten im verwendeten Datensatz erfolgreicher wäre, da bei dieser Methode weniger Bewertungen zur erfolgreichen Manipulation notwendig sind und sie öfters erfolgreich ist (weniger als 100 Bewertungen zur Manipulation notwendig).

In ihrem Konferenzpapier stellen Banerjee und Chua (2014) ein Experiment dazu vor, wie gefälschte Bewertungen von Hotels verfasst werden. Bei diesem wurden Probanden mit dem Verfassen gefälschter Bewertungen beauftragt. Anschließend sollten sie einen qualitativen Fragebogen zu ihrem Vorgehen ausfüllen. Ergebnis ist, dass die Probanden zum Verfassen ihrer Bewertungen zunächst andere Bewertungen gelesen haben sowie Google als Suchmaschine genutzt haben um Informationen zu einem Hotel zu erlangen. Das Verfassen der Bewertungen wurde als nicht besonders herausfordernd, jedoch zeitaufwendig beschrieben.

Ein Experiment, welches dem dieser Arbeit ähnlich ist, wurde 2010 und 2012 von der Stiftung Warentest bezogen auf Hotel-Bewertungsplattformen durchgeführt. Neben diversen anderen Bewertungskriterien wurde untersucht, wie die getesteten Plattformen mit gefälschten Bewertungen umgehen. In die Bewertung einbezogen wurde dabei nicht nur die Löschquote der

abgegebenen Bewertungen, sondern auch ob der Autor einer Bewertung über die Löschung informiert wurde. Bei jedem der dreizehn untersuchten Portale wurden dafür (sofern möglich) fünf Bewertungen verfasst. Diese enthielten folgende Kriterien: „Werbeaussagen, Blindtext, auffällige falsche Angaben zum Standort“ (test, 2012, S. 80). Im ähnlich aufgebauten Experiment 2010 wurden darüber hinaus Bewertungen abgegeben, welche die „Bewertung eines Hotelbereichs [betrafen], den das Hotel nicht besitzt“ (test, 2010, S. 79). Diese Kriterien wurden für das Experiment dieser Arbeit aufgegriffen und leicht abgewandelt ebenfalls angewandt (siehe Kapitel 3.3.2). Im Experiment der Stiftung Warentest wurde die Erkennung gefälschter Bewertungen der Plattformen zweimal mit „gut“, je dreimal mit „befriedigend“ sowie „ausreichend“ und viermal mit „mangelhaft“ bewertet. „Mangelhaft“ wurde Google bewertet, da dort alle gefälschten Bewertungen freigeschaltet wurden. Google Maps ist die einzige Plattform im Test der Stiftung Warentest, welche auch in dieser Arbeit untersucht wurde, da es sich nicht um eine reine Hotelbewertungsplattform handelt.

2.5 Vorstellung der Plattformen

Für das im Rahmen dieser Arbeit durchgeführte Experiment wurden vier Plattformen ausgewählt, auf denen Nutzer Bewertungen für Artikel bzw. für Orte abgeben können. Dabei handelt es sich um Facebook, Google Maps, Google Play Store und Zalando. Warum diese Plattformen ausgewählt wurden, ist in Kapitel 3.1 begründet. Im Folgenden werden die Plattformen im Fokus ihrer Bewertungsfunktion vorgestellt.

2.5.1 Facebook

Mit über zwei Milliarden aktiven Nutzern im Monat ist Facebook das weltweit meistbesuchte soziale Netzwerk (Facebook, 2017a). Nutzer können auf Facebook Profile erstellen und vielfältig mit anderen Nutzern interagieren. Neben Profilen für Personen können Nutzer zudem Seiten für Unternehmen, Prominente oder Orte erstellen und pflegen. Der Administrator einer Facebook-Seite kann einstellen, dass alle anderen Nutzer die Seite bewerten können. Ist in der Seite ein Shop integriert, wird die Bewertungsfunktion von Facebook zwangsläufig aktiviert (Facebook, 2017c). Die Bewertung erfolgt auf einer Skala von einem bis fünf Sternen sowie einem optionalen Freitext (40 bis 10.000 Zeichen). Die Nutzer können die Sichtbarkeit einer Bewertung einschränken und diese beispielsweise nur für Freunde freigeben. In die Durchschnittsbewertung einer Seite gehen allerdings nur die Bewertungen ein, die alle Facebook-Nutzer einsehen können, also als „öffentlich“ abgegeben wurden (Facebook, 2017d). Die durchschnittliche Bewertung wird auf Seiten in der rechten Seitenleiste als oberstes Element angezeigt. Alle Bewertungen sind für die Besucher einer Seite auf einer gesonderten Unterseite einsehbar. Die voreingestellte Sortierung der Bewertungen ist „Am Hilfreichsten“. Nach welchen

Kriterien diese Sortierung berechnet wird, gibt Facebook nicht bekannt. Alternativ können die Bewertungen nach Zeitpunkt („Neueste“) und nach Anzahl der vergebenen Sterne sortiert werden. Außer auf der Seite selbst werden Bewertungen von Seiten etwa auch in den Suchtreffern unterhalb des Namens der Seite angezeigt. Der Nutzer kann sie somit in seine Entscheidung einbeziehen, ob er eine Seite anklickt.

2.5.2 Google Maps

Bei Googles Kartendienst Maps können dort eingetragene Orte durch Nutzer bewertet werden. Dafür muss der Nutzer über einen Google-Account verfügen. Die Bewertung erfolgt wie bei Facebook über einen bis fünf Sterne sowie einen optionalen Freitext (1 bis 4096 Zeichen). Für alle Nutzer einsehbar sind die Bewertungen beim Klick auf einen Ort auf der Landkarte. In der Detailansicht zu einem Ort ist zunächst die durchschnittliche Sterne-Bewertung direkt unter dem Namen eines Ortes zu sehen. Durch Scrollen sieht der Nutzer außerdem eine Übersicht über die Anzahl der Bewertungen pro Sterne und drei Bewertungen, die von Google ausgewählt wurden. Diese sind auf etwa 50 bis 100 Zeichen (in der Regel ein Satz) gekürzt und einige Wörter sind hervorgehoben. Durch einen zusätzlichen Klick kann der Nutzer zudem alle Rezensionen einsehen. Diese sind zunächst nach „Hilfreichste“ sortiert, können aber auch zeitlich („Neueste“) oder nach Anzahl der Sterne („Höchste/Niedrigste Bewertung“) sortiert werden. Darüber hinaus wird die durchschnittliche Bewertung eines Ortes auch in den Suchergebnissen von Google Maps sowie in der Google-Websuche angezeigt. Besonders durch die Integration in der Websuche („Google Knowledge Graph“) können Bewertungen potenziell an eine große Anzahl von Nutzern ausgespielt werden, die diese ggf. zur Meinungsbildung über einen Ort verwenden.

2.5.3 Google Play Store

Unter dem Namen „Google Play“ wurden 2012 der Android Market (Apps), Google Music und Google eBookstore zusammengefasst (Official Google Blog, 2012). Über die gleichnamige Android-App oder einen Webbrowser können Nutzer mit Google-Account Android-Apps, Musik, Filme und Serien sowie Bücher und Zeitschriften als Downloads erwerben. Alle Artikel im Google Play Store können von angemeldeten Nutzern bewertet werden. Auch hier gibt es ein System mit maximal fünf Sternen und einem optionalen Freitext mit maximal 1.200 Zeichen. Die Darstellung der Bewertungen ähnelt ebenfalls der von Google Maps. Unter dem Titel eines Artikels wird die durchschnittliche Bewertung angezeigt, weiter unten werden neben einer Übersicht über die Sterne-Verteilung drei bis vier Bewertungen angezeigt. Diese sind im Gegensatz zu Google Maps allerdings nicht gekürzt. Alle weiteren Bewertungen sind über einen Klick aufrufbar. Die Sortierung der Bewertungen ist für den Nutzer nur bei Apps veränderbar. Bei Musik, Filmen und Büchern richtet sie sich nicht (allein) nach Datum oder Sterneanzahl und

ist mit der Sortierung „Am Hilfreichsten“ anderer Plattformen vergleichbar. Die Bewertung von Artikeln im Google Play Store ist nur auf Geräten möglich, auf denen diese auch genutzt werden können. Apps können daher nur von Android-Geräten und nicht in der Web-Oberfläche bewertet werden. Da im Experiment dieser Arbeit nicht genügend Endgeräte für Bewertungen von Apps zur Verfügung standen, wurden ausschließlich Filme, Bücher und Musik bewertet.

2.5.4 Zalando

Nach Erhebungen des EHI Retail Institute (2017) ist Zalando mit rund 1,2 Milliarden Euro Umsatz im Jahr 2016 nach Amazon und OTTO der drittgrößte Online-Händler in Deutschland. Kunden können dort Mode, Schuhe und Accessoires erwerben. Alle Artikel können von angemeldeten Nutzern bewertet werden. Die Bewertung erfolgt mit einem bis fünf Sternen sowie einem optionalen Text (maximal 1500 Zeichen). Voraussetzungen für das Verfassen von Bewertungen wie etwa einen Mindestumsatz gibt es bei Zalando nicht, der Nutzer muss lediglich über einen Account bei Zalando verfügen. Auf der Produktseite werden die Bewertungen bei Zalando an zwei Stellen angezeigt. Ohne zu scrollen sieht der Nutzer die durchschnittliche Sterne-Bewertung eines Produktes, die unter dem Namen und dem Preis eines Produktes rechts neben dem Produktbild angezeigt wird. Alle Bewertungen inklusive der dazugehörigen Texte kann der Nutzer einsehen, indem er je nach Bildschirmgröße etwas nach unten scrollt und anschließend auf „Bewertungen ([Anzahl der Bewertungen])“ klickt. An der Stelle, an der zuvor weitere Informationen zum Produkt zu finden waren, erscheinen dann die Bewertungen sortiert nach Erstellungsdatum (neueste zuerst). Damit sind die Bewertungen etwas weniger prominent auf der Artikelseite präsentiert als bei vielen anderen Online-Shops, wo der Nutzer nur nach unten scrollen muss um Bewertungen lesen zu können. Auch in den Suchergebnissen oder auf Kategorie-Seiten werden keine Bewertungen angezeigt.

2.6 Regeln der Plattformen gegen gefälschte Bewertungen

Google und Facebook stellen für ihre Plattformen Richtlinien zur Verfügung, an die die Nutzer sich beim Verfassen von Bewertungen halten sollen. Bei Google gibt es getrennte Richtlinien für Maps und Play Store, die sich jedoch überschneiden. Zalando stellt in seinem Hilfebereich keine Informationen zu Bewertungen bereit und formuliert auch an anderer Stelle keine öffentlichen Anforderungen an diese.

2.6.1 „Gemeinschaftsstandards“ bei Facebook

In Facebooks Hilfebereich (Stand 15.01.2018) wird erläutert, dass Bewertungen persönliche Erfahrungen zugrunde liegen müssen und dass sie sich auf „das Produkt oder die Dienstleistung beziehen [müssen], das bzw. die von der Seite angeboten wird“ (Facebook, 2017e). Zudem wird auf die Facebook Gemeinschaftsstandards für die gesamte Plattform verwiesen, die unter den Aspekten „Unterstützung für deine Sicherheit“, „Respektvollen Umgang fördern“, „Deine Kontoinformationen und persönlichen Daten schützen“ und „Dein geistiges Eigentum schützen“ stehen (Facebook, 2017b). Facebook schreibt, dass Bewertungen, die die aufgestellten Regeln verletzen, *möglicherweise* gelöscht werden. Die Plattform scheint sich somit im Klaren darüber zu sein, dass nicht konforme Bewertungen nicht in jedem Fall gelöscht werden oder gelöscht werden können.

2.6.2 „Richtlinien für Rezensionen“ bei Google Maps

Auch Google Maps informiert den Nutzer über eine Löschung „unter Umständen“, wenn eine Bewertung nicht den aufgestellten Regeln entspricht (Google My Business-Hilfe, 2017). Darüber hinaus sind die Richtlinien für Rezensionen bei Google Maps (Stand 15.01.2018) die umfangreichsten unter den untersuchten Plattformen, die sich explizit auf Rezensionen beziehen. Kern der Richtlinien ist es, die Authentizität der Bewertungen zu garantieren. So sollen etwa keine Bewertungen verfasst werden, wenn es Interessenskonflikte gibt und es dürfen keine Bewertungen mit mehreren Konten erstellt werden. Auch muss sich eine Bewertung bei Google Maps auf den bewerteten Ort beziehen und darf keine Werbung enthalten.

2.6.3 „Richtlinien zum Posten von Kommentaren“ im Google Play Store

Bei Google Play wird zwischen „Richtlinien“, deren Einhaltung vom Nutzer erwartet wird, und „Richtlinien und Tipps“ (beides in den Versionen vom 15.01.2018 betrachtet) unterschieden, um deren Einhaltung der Nutzer gebeten wird (Google Play, 2017). Während die „Richtlinien und Tipps“ den Nutzer dazu anhalten sollen einen gut lesbaren und hilfreichen Text abzugeben, ähneln die „Richtlinien“ inhaltlich denen von Google Maps. So muss sich eine Bewertung auf die bewertete App beziehen und es dürfen keine Links oder Codes für Partnerprogramme enthalten sein. Auf Bewertungen von Medien wird in den Richtlinien kein gesonderter Bezug genommen. Darüber hinaus werden wie bei Google Maps und Facebook etwa Hassbotschaften und obszöne Inhalte verboten. Explizit erwähnt wird auch das Verbot gefälschter Bewertungen zur Verzerrung der Gesamtbewertung.

Zusammenfassend lässt sich festhalten, dass die Plattformen mit Ausnahme von Zalando umfangreiche Richtlinien für Bewertungen auf ihren Seiten haben. Diese müssen von den Nutzern allerdings selbst im Hilfebereich gesucht werden und werden nicht vor dem Verfassen

einer Rezension angezeigt. Die von Google aufgestellten Regeln sind dabei etwas umfassender als die von Facebook, aber auch dort reichen die Regeln um authentische und möglichst hilfreiche Bewertungen einzufordern. Da sich nicht alle Nutzer an die aufgestellten Regeln halten werden, stellt sich die Frage, wie gut die Plattformen diese Regeln mit Löschungen durchsetzen.

Im folgenden Kapitel wird erläutert, wie das Experiment dieser Arbeit durchgeführt wurde, mit dem diese Frage beantwortet wurde.

3 Methodik und Vorgehensweise

Um möglichst aussagekräftige Ergebnisse zu erhalten, wurden sämtliche im Experiment abgegebene Bewertungen systematisch und mit vorher festgelegten Inhalten verfasst. Wesentlicher Teil dieser Arbeit war so die Planung der Inhalte und Bedingungen, unter denen die Bewertungen abgegeben wurden.

3.1 Auswahl der Plattformen

Wie bereits in der Motivation erläutert, sind durch Nutzer verfasste Bewertungen im Internet heute allgegenwärtig. Dementsprechend groß war die Auswahl der Bewertungsplattformen bzw. Webseiten mit Bewertungsfunktion, die in das Experiment dieser Arbeit hätten einbezogen werden können. Aufgrund ihrer Beliebtheit wurden letztlich Facebook, Google Maps und der Google Play Store ausgewählt. Ebenfalls unter dem Kriterium der großen Nutzerzahl wurde Zalando ausgewählt. Zusätzlicher Grund für die Auswahl von Zalando war jedoch auch, dass es der größte Online-Shop in Deutschland ist (EHI Retail Institute, 2017), der für das Verfassen einer Bewertung keinen Mindestumsatz fordert. Bei Amazon muss ein Kunde beispielsweise mindestens 50€ Umsatz generiert haben, bevor das Bewerten von Artikeln für ihn freigeschaltet ist. Auch OTTO erlaubt Nutzern erst nach ihrer ersten Bestellung das Verfassen von Bewertungen. Die Durchführung des Experiments im Rahmen einer Bachelorarbeit war damit auf diesen Plattformen nicht möglich. Entscheidend für die Auswahl von großen Plattformen war zudem der Anspruch, die Folgen durch die abgegebenen Bewertungen für die Plattformen möglichst gering zu halten. Potenziell kostet die Entfernung von gefälschten Bewertungen Ressourcen und diese sind für große Konzerne wesentlich leichter aufzubringen als für kleinere Online-Shops oder Bewertungsportale.

3.2 Kontrolle der Folgen der abgegebenen Bewertungen

Teilziel des Experiments war es, bei dessen Durchführung keiner der von den abgegebenen Bewertungen betroffener Parteien einen nennenswerten Schaden zuzufügen. Neben der Auswahl großer Plattformen wurden daher auch die dort bewerteten Artikel/Orte mit Bedacht ausgewählt. Grundvoraussetzung für die Auswahl eines Bewertungsgegenstandes war, dass dieser bereits über eine mindestens zweistellige Anzahl an Bewertungen verfügte. Bei den meisten ausgewählten Artikeln/Orten lag die Anzahl der schon vorhandenen Bewertungen im mittleren zweistelligen bis niedrigen dreistelligen Bereich, teilweise aber auch deutlich darüber. So fielen die im Rahmen des Experiments abgegebenen Bewertungen wenig bis nahezu gar nicht ins Gewicht der Durchschnittsbewertung.

Ebenfalls wurde darauf geachtet, dass durch die verfassten Bewertungen keine Nutzer über die Qualität eines Produktes oder Ortes getäuscht wurden. Keine der im Rahmen des Experiments abgegebenen Bewertungen enthält einen Freitext, der vermeintlich qualifiziert das Produkt bewertet.

Bei der Auswahl der Orte auf Facebook und Google Maps wurde darauf geachtet, dass keine Einzelhändler, Hotels, Restaurants, Ärzte oder andere Firmen bewertet wurden, bei denen eine Bewertung bereits Einfluss auf die Umsätze haben kann. In erster Linie wurden Sehenswürdigkeiten und andere öffentliche Orte bewertet, bei denen es keine direkten und höchstens minimale potenzielle Auswirkungen durch die abgegebenen Bewertungen geben kann. Zudem wurden alle Konten und Bewertungen nach Ende des Experiments gelöscht.

3.3 Klassifikation der Konten und Bewertungen

Alle im Rahmen des Experiments angelegten Konten und verfassten Bewertungen wurden nach zuvor festgelegten Kriterien erstellt. Dies war eine Grundvoraussetzung um systematisch feststellen zu können, wie die Plattformen bei der Freischaltung von Bewertungen vorgehen. Ähnliche Kriterien wurden in der Vorbereitung zu Klassen zusammengefasst. Die entstandenen Klassen der Kontoerstellung konnte so mit den Klassen der Bewertungen je zweimal kombiniert werden ohne einen zu großen Aufwand durch zu viele Kriterien zu erzeugen.

3.3.1 Konten

Die Konten bei den vier Plattformen wurden je nach den Kriterien von einer der drei definierten Klassen erstellt. Ziel der drei Klassen war es, eine Abstufung im Faktor der Plausibilität, dass sich hinter dem Konto ein echter Nutzer verbirgt, zu erzeugen. Für die erste Klasse (**Wenig verdächtig**) bedeutete dies, dass die Accounts bei Abgabe der ersten Bewertung mindestens eine Woche alt waren. Zudem wurde nach Erstellung des Accounts und bei Abgabe der Bewertung zusätzlich plattformtypische Interaktion simuliert. Mit Google-Accounts wurden etwa E-Mails geschrieben, YouTube-Videos aufgerufen oder Routen auf Google Maps recherchiert. Bei Facebook-Accounts wurden ein Profilbild eingerichtet und weitere Angaben im Profil ausgefüllt. Darüber hinaus haben die Facebook-Accounts, die für diese Klasse erstellt wurden, Freundschaften geknüpft, Nachrichten ausgetauscht, Seiten abonniert und die Mitgliedschaft bei Gruppen angefragt. Bei Zalando wurde in Produktkategorien gestöbert und Artikel wurden zur Wunschliste oder in den Warenkorb hinzugefügt.

Die Accounts der Klasse 2 (**Verdächtig**) waren zum Zeitpunkt der Bewertungsabgabe einen Tag alt. Bei ihnen wurden neben der Bewertung ein oder zwei plattformtypische Interaktionen durchgeführt und beispielsweise ein Profilbild festgelegt oder einige Artikel aufgerufen.

Accounts der Klasse 3 (**Sehr verdächtig**) wurden offensichtlich allein zum Erstellen einer/mehrerer Bewertungen angelegt. Abgesehen von der Kontenerstellung und der unmittelbar danach durchgeführten Bewertung wurden keine weiteren Aktionen ausgeführt.

Tabelle 1: Klassen der Konten

Klasse 1 – Wenig verdächtig	Klasse 2 – Verdächtig	Klasse 3 – Sehr verdächtig
<ul style="list-style-type: none"> • Alter: Älter als eine Woche • Plattformtypische Interaktionen (z. B. liken, suchen, Nachrichten schreiben) 	<ul style="list-style-type: none"> • Alter: Einen Tag • Minimale, plattformtypische Interaktion 	<ul style="list-style-type: none"> • Bewertung sofort nach der Erstellung des Kontos veröffentlicht • Keine andere Interaktion

3.3.2 Bewertungen

Die Bewertungen wurden nach den Regeln von vier verschiedenen Klassen erstellt. Zwei davon hatten den Schwerpunkt „Spam“ und zwei den Schwerpunkt „Betrug“, jeweils abgestuft in der geschätzten Schwierigkeit der Erkennung für die Plattformen. Im Folgenden werden diese Klassifizierungen vorgestellt und es wird erläutert, inwieweit die enthaltenen Kriterien die Richtlinien der Plattformen verletzen (sofern Richtlinien vorhanden sind, siehe Kapitel 2.6). Grundsätzlich ist anzumerken, dass alle abgegebenen Bewertungen die Richtlinien von Google und Facebook allein dadurch verletzten, dass keine von ihnen die tatsächliche Meinung des Autors zu einem Bewertungsgegenstand enthält.

Bewertungen der Klasse „**Betrug 1**“ sollten die schnelle Erzeugung von Bewertungen zur Verbesserung des Gesamtbildes der Bewertungen eines Artikels/Ortes abbilden. Daher wurden bei allen Bewertungen dieser Klasse vier oder fünf Sterne vergeben. Der Text der Bewertung enthielt dabei keine Bewertung, sondern in „Werbesprache“ formulierte Beschreibungen oder aus der Detailansicht eines Bewertungsgegenstandes kopierte Informationen wie Name, Adresse oder Produkteigenschaften. Die Bewertungen basieren damit nicht auf einer persönlichen Erfahrung und widersprechen so den Richtlinien der Plattformen.

Beispiele für Bewertungen der Klasse Betrug 1:

- *Der Alexanderplatz ist Berlins meistbesuchte Destination – hier bewegen sich täglich mehr als 360.000 Menschen.*
- *Essener Dom Kettwiger Str. 42, 45127 Essen*
- *The Blues Brothers: Original Soundtrack Recording 1980 Atlantic Recording Corp. Marketed by Rhino Entertainment Company, a Warner Music Group Company.*

- *Längenverstellbare Schultergurte, Klett- sowie Reißverschlussfach innen, Reißverschlussfach hinten, zusätzlicher, längenverstellbarer Karabinerverschluss vorne*
- *Die nächste Party kommt bestimmt! Mit diesem Cocktailkleid machst Du garantiert eine gute Figur!*

Bei Bewertungen der Klasse „**Betrug 2**“ wurden ausschließlich Freitexte mit inhaltlichen Fehlern abgegeben. Beispielsweise wurden Eigenschaften eines Ortes gelobt oder kritisiert, über die dieser offensichtlich nicht verfügt. Auch wurden Bewertungen abgegeben, die sich nicht auf das Produkt beziehen, sondern beispielsweise den Versand oder die Downloadgeschwindigkeit bewerten. Dabei wurden gleich viele positive (4-5 Sterne) und negative (1-2 Sterne) Bewertungen verfasst. Real treten die Kriterien dieser Klasse beispielsweise auf, wenn der Auftragsautor einer bezahlten Bewertung den Bewertungsgegenstand nicht kennt. Auch hier werden die Richtlinien von Facebook und Google verletzt, da die Rezensionen nicht auf persönlichen Erfahrungen basieren und sich teilweise nicht auf den Bewertungsgegenstand beziehen.

Beispiele für Bewertungen der Klasse Betrug 2:

- *Der Eintritt ist viel zu hoch! (Bei einer frei zugänglichen Sehenswürdigkeit)*
- *Schönes a cappella Album! (Es handelte sich nicht um ein a cappella Album)*
- *Download ging sehr schnell*
- *Versandzeit zwei Tage, das hätte schneller gehen können.*
- *Den Karton finde ich viel zu groß für einen Gürtel*

Für die Klasse „**Spam 1**“ wurden Bewertungen abgegeben, die aus Blindtexten bestehen. Dabei handelte es sich zum Beispiel um das im Webdesign verbreitete „Lorem Ipsum“ oder um Zitate aus der Wikipedia oder deutschsprachigen Literaturklassikern. Zudem wurden in dieser Kategorie Bewertungen in sehr schlechtem Sprachstil abgegeben, beispielsweise mit groben Rechtschreib-/Tippfehlern, willkürlicher Zeichensetzung oder ausschließlicher Verwendung von Großbuchstaben. Während dies bezogen auf die Richtlinien der Plattformen grenzwertig ist, verletzen Blindtexte die Regeln klar, da es sich um Spam handelt, dem jeglicher Bezug zum Bewertungsgegenstand fehlt. In dieser Klasse wurden ebenfalls gleich viele positive (4-5 Sterne) und negative (1-2 Sterne) Bewertungen abgegeben.

Beispiele für Bewertungen der Klasse Spam 1:

- *Als Gregor Samsa eines Morgens aus unruhigen Träumen erwachte, fand er sich in seinem Bett zu einem ungeheueren Ungeziefer verwandelt. Und es war ihnen wie eine*

Bestätigung ihrer neuen Träume und guten Absichten, als am Ziele ihrer Fahrt die Tochter als (Zitat aus Franz Kafkas „Die Verwandlung“)

- *Ein Hyperlink (englische Aussprache [ˈhaɪpeɪlɪŋk], deutsch wörtlich „Über-Verknüpfung“, sinngemäß elektronischer Verweis), kurz Link, ist ein Querverweis in einem Hypertext, der funktional einen Sprung zu einem anderen elektronischen Dokument oder an eine* (Zitat aus der Wikipedia)
- *Vol der kirrkte Plats kan Mann geil dschilhen !!!! Im somer*
- *FILMHATMIRNICHTSOGUTGEFALLENNICHTSOSPANNEND*

Wie bereits erwähnt, verlangen Facebook und Google in ihren Richtlinien, dass einer Bewertung tatsächliche eigene Erfahrungen zugrunde liegen müssen. Daher wurden in der Klasse „**Spam 2**“ Bewertungen verfasst, in denen der Autor klar zu erkennen gibt, dass er das Produkt/den Ort nicht kennt und seine Meinung nur anhand von Aussagen Dritter oder basierend auf Vermutungen gebildet hat. Dies kann zwar eine Meinung zu einem Bewertungsgegenstand sein, diese ist jedoch wesentlich weniger aussagekräftig als eine Bewertung basierend auf eigenen Erfahrungen und widerspricht in jedem Fall den Plattformrichtlinien von Facebook und Google. Auch in der Klasse „Spam 2“ wurden wie in den beiden vorhergehenden Klassen positive und negative Bewertungen verfasst.

Beispiele für Bewertungen der Klasse Spam 2:

- *Ich würde gerne mal zum Kölner Dom aber habe es leider selbst noch nicht geschafft.*
- *Ich habe ein Doku gesehen voll viel Müll in Hamburg besser nicht hin :-)*
- *Leider noch nicht angesehen (bewertet wurde ein Film)*
- *Auf dem Weihnachtswunschzettel für dieses Jahr!*
- *Ich mag oversized Sachen nicht so, deshalb nicht bestellt*

In Tabelle 2 sind die vier Klassen für Bewertungen in einer Tabelle zusammengefasst:

Tabelle 2: Klassen der Bewertungen

Betrug 1	Betrug 2
Bewertung enthält nur kopierten Text aus Produktinformationen/Beschreibung des Ortes oder Werbesprache	Bewertung bezieht sich nicht auf Produkt/Ort, etwa durch inhaltsleere Aussagen oder durch Bewertung nicht vorhandener Eigenschaften
Spam 1	Spam 2
Blindtexte oder sehr schlecht lesbare Texte in Versalien mit vielen Fehlern	Bewertung enthält die eindeutige Aussage, dass der Autor den Bewertungsgegenstand nicht kennt.

3.4 Erstellung der Konten

Für das Experiment wurden je 24 Konten pro Dienst erstellt, was insgesamt 96 Konten (davon 48 bei Google) entspricht. Dabei wurde zunächst ein Google-Konto erstellt und dieses in 50% der Fälle anschließend zur Erstellung eines Facebook- und Zalando-Kontos genutzt. Um ein Facebook-Konto nutzen zu können, muss ein an die angegebene E-Mail-Adresse gesendeter Code eingegeben werden oder ein „Login mit Google“ nach dem eigentlichen Anmeldeprozess erfolgen. Wurde ein Facebook-Konto so über eine E-Mail-Adresse von Google Mail bestätigt, ist es uneingeschränkt nutzbar. Eine Verifizierung per SMS ist optional.

Bei Zalando wird im Rahmen des Registrierungsprozesses keine Verifikation der angegebenen E-Mail-Adresse durchgeführt. Ob dies später bei einer Bestellung von Waren erfolgt, wurde in dieser Arbeit nicht untersucht. Grundsätzlich kann ein Nutzer ein Konto mit einer fiktiven E-Mail-Adresse erstellen und mit diesem etwa einen Wunschzettel füllen oder Bewertungen abgeben. Im Rahmen der Registrierung sendet Zalando jedoch eine „Willkommens-Mail“ an den neuen Nutzer. Diese E-Mail wurde im Experiment bei den Konten der Klassen „Wenig verdächtig“ immer und bei den Konten der Klasse „Verdächtig“ teilweise geöffnet. Aufgrund der enthaltenen HTML-Inhalte mit möglicherweise individuellen Grafiken („Zählpixel“) in der E-Mail besteht technisch die Möglichkeit, dass Zalando die E-Mail-Adresse nach dem Öffnen der E-Mail als verifiziert einstuft.

Die Erstellung eines Google-Kontos ist prinzipiell ohne die Angabe einer anderen E-Mail-Adresse oder einer Telefonnummer möglich. Google führt jedoch im Registrierungsprozess eine Abschätzung durch, wie hoch die Wahrscheinlichkeit ist, dass es sich nicht um eine legitime Kontenerstellung durch einen echten Nutzer handelt. Hält Google diese Wahrscheinlichkeit für zu hoch, ist die Verifikation des Kontos per SMS-Bestätigungscode oder über einen Sprachanruf

mit vorgelesenem Bestätigungscode unumgänglich um das Konto nutzen zu können. Näheres dazu ist in Kapitel 4.3 beschrieben. Die im Rahmen des Experiments erstellten Konten wurden ohne Verifizierung angelegt. Dafür wurden stets private Internetanschlüsse verwendet. Die IP-Adressen wurden etwa durch das Neustarten des Routers oder durch den gewöhnlichen IP-Adressen-Wechsel der Provider zwischen dem Erstellen der Konten gewechselt. Nicht gewechselt werden konnten die Adressräume. Da Anschlüsse verschiedener Provider (DSL, Kabelnetz, Mobilfunk) an unterschiedlichen Orten genutzt wurden, konnte jedoch auch bei den Adressräumen etwas Diversität geschaffen werden. Jenseits der verschiedenen IP-Adressen wurden verschiedene Browser verwendet (siehe Kapitel 3.5) und nicht mehr als fünf Google-Konten mit einem Endgerät an einem Tag erstellt. Nicht gewechselt wurde bei allen Registrierungen über Desktop-Browser das Betriebssystem und die Hardware. Die Plattformen haben so die Möglichkeit gehabt, Hinweise auf den gleichen Ursprung für viele Konten zu erhalten.

3.5 Browserumgebungen

Um zu simulieren, dass stets neue legitime Benutzer eine Account-Erstellung durchführen, wurde für jedes Konto eine andere Browserinstanz verwendet. Dies wurde erreicht, indem die portablen Versionen von Google Chrome, Mozilla Firefox und Opera für das Experiment verwendet wurden. Um die 48 Instanzen weiter zu diversifizieren wurde bei einigen die Standard-Schriftart, -Schriftgröße, -Fenstergröße und die Sprache („Deutsch“/„Deutsch (Deutschland)“) verändert. Einige Konten wurden zudem auf Smartphones und Tablets angelegt. Hier wurden ebenfalls verschiedene Browser genutzt sowie vor und nach der Kontenerstellung die Cookies und Offline-Daten gelöscht um eine Beeinflussung durch das echte Surfverhalten der Geräte-Besitzer zu verhindern. Eventuelle weitere Logins in das Konto erfolgten in diesen Fällen in der Regel über einen Desktop-Browser. Dort konnten einige Faktoren, die das Wiedererkennen eines Benutzers erleichtern, nicht geändert werden. Beispielsweise das Betriebssystem, die IP-Adressräume und das Gerät, mit dem die Konten erstellt wurden. Letzteres kann in vielen Fällen durch das Fingerprinting von HTML Canvas-Elementen wiedererkannt werden (Mowery und Shacham, 2012, S. 1). Die Anmeldeseiten von Facebook und Zalando enthalten Canvas-Elemente, diese müssen jedoch nicht zwangsläufig zur Erkennung von Nutzern eingesetzt werden.

Durch die beschriebenen Umgebungen gab es für die Plattformen Hinweise, dass eine große Zahl von Bewertungen tatsächlich von nur einer Person abgegeben wurde, eindeutig war dies aber nicht. Durch die Arbeit mit vielen Browser-Instanzen wurde zudem gewährleistet, dass die Webseiten (etwa im Gegensatz zum Surfen im privaten Modus) Cookies und andere Offline-

Daten anlegen konnten. Das Vorhandensein von Cookies aus vergangenen Logins erhöht die Authentizität von Konten bei wechselnden IP-Adressen.

In Tests vor der eigentlichen Durchführung des Experiments wurde erprobt, Konten über das Tor-Netzwerk zu erstellen. Dies sollte ursprünglich ein Faktor für Konten der Klasse „Sehr verdächtig“ sein. Die Tests haben allerdings gezeigt, dass eine Kontoerstellung bei Zugriff über das Tor-Netzwerk bei Google nicht ohne Verifizierung einer Telefonnummer möglich ist und bei Facebook ein Foto des eigenen Gesichts hochgeladen werden muss. Dieses Verfahren wurde von Facebook erst kurz zuvor im Herbst 2017 eingeführt und soll die Authentizität eines Kontos mit verdächtigen Aktivitäten sicherstellen (Tiku, 2017). Lediglich bei Zalando war das Erstellen von Konten genauso wie über einen normalen Internetanschluss möglich. Da die Nutzung von Tor nur bei Zalando nicht sinnvoll gewesen wäre, wurde im Experiment vollständig auf die Erstellung von Konten über das Tor-Netzwerk verzichtet.

3.6 Erstellung der Bewertungen

Grundsätzlich wurden alle Bewertungen bereits vor der Durchführungsphase des Experiments vorbereitet. Dies hat die im Versuchszeitraum in der Regel täglich stattfindende Durchführungsroutine des Experiments vereinfacht. Da auf jeder Plattform andere Artikel/Orte bewertet wurden, wurden Bewertungen in der Regel nur einmal abgegeben. Allein einige Bewertungen wie Blindtexte ohne Bezug zum Bewertungsgegenstand konnten auf mehreren Plattformen abgegeben werden und der Umgang damit somit eindeutig miteinander verglichen werden.

Jeweils die Hälfte der angelegten Konten pro Plattform hat im Rahmen des Experiments nur eine Bewertung abgegeben. Über die andere Hälfte der Konten wurden jeweils drei Bewertungen für drei verschiedene Bewertungsgegenstände verfasst. Jede Kombination aus Konten-Klasse und Bewertungs-Klasse wurde pro Plattform zweimal umgesetzt. Einmal mit einem Account, über den eine Bewertung veröffentlicht wurde und einmal mit einem Account über den drei Bewertungen veröffentlicht wurden. Dies entspricht vier Bewertungen pro Klassen-Kombination auf einer Plattform, 48 Bewertungen pro Plattform und 192 Bewertungen insgesamt. Der Zeitpunkt der Abgabe einer Bewertung wurde vor der Durchführung des Experiments festgelegt und war abhängig von der Klasse des zugehörigen Kontos.

Bei Konten der Klasse „Wenig verdächtig“ wurde die erste Bewertung acht bis zehn Tage nach dem Tag der Kontenerstellung abgegeben. Wenn ein Konto der Klasse „Wenig verdächtig“ drei Bewertungen abgeben sollte, wurde in einigen Fällen zudem zwischen der ersten und zweiten oder der zweiten und dritten Bewertung eine eintägige Pause eingelegt. Verbunden mit den um

die Bewertungsabgaben durchgeführten plattformtypischen Interaktionen wurde so der Anschein, dass sich ein legitimer Nutzer hinter einem Konto verbirgt, weiter gestärkt.

Eine Wartezeit von ein bis zwei Tagen zwischen der Kontenerstellung und Bewertungsabgabe wurde bei Konten der Klasse „Verdächtig“ eingehalten. Mehrere Bewertungen wurden in dieser Kategorie mit Pausen von maximal wenigen Minuten hintereinander abgegeben. Auf die Konten wurden in der Durchführungsphase des Experiments somit nur zweimal zugegriffen: Bei der Erstellung und beim Abgeben der Bewertung bzw. Bewertungen.

Bei Konten der Klasse „Sehr verdächtig“ erfolgte die Abgabe der Bewertung unmittelbar nach Abschluss der Kontenerstellung und somit als einzige Interaktion. Konten dieser Klasse haben daher im Versuchszeitraum nur eine einzige Session zu verzeichnen (zzgl. eines weiteren Logins zur Kontolöschung nach Ende des Experiments).

Anders als die Kontenerstellung erfolgte die Abgabe von Bewertungen (sofern es sich dabei um getrennte Sessions handelte) auch über (halb-)öffentliche Internetzugänge in Bahnhöfen und Hochschulen. Da Google die Abwägung zur Notwendigkeit einer Telefonnummerverifikation nur bei der Registrierung trifft und zudem in der Regel bereits Cookies von den Plattformen in den jeweiligen Browser-Instanzen gesetzt wurden, stellte dies kein Problem dar. Das Nutzen von öffentlichen Internetzugängen widerspricht nicht dem typischen Verhalten echter Nutzer und so konnte auf diesem Weg die Anzahl der verschiedenen IP-Adressräume, von denen auf die verschiedenen Konten zugegriffen wurde, erhöht werden.

3.7 Kontrolle des Online-Status

Nach dem Abgeben einer Bewertung wurde deren Online-Status regelmäßig kontrolliert. Dazu wurde der Bewertungsgegenstand in einem privaten Browserfenster aufgerufen und nach der für dieses Experiment abgegebenen Bewertung gesucht. Die Kontrolle der Sichtbarkeit einer Bewertung für alle Nutzer erfolgte nach einem festen Rhythmus: 20 bis 30 Stunden nach der Abgabe, drei Tage nach der Abgabe, sieben Tage nach der Abgabe und final fünfzehn Tage nach der Abgabe. Protokolliert wurde jeweils der letzte Status einer Bewertung (online/nicht online) sowie im Falle einer Veränderung dieses Status der letzte gemessene Zeitpunkt des vorherigen Status.

3.8 Reflexion der Methodik

Nach Maßstäben anderer Disziplinen wie den Sozialwissenschaften oder der Psychologie handelt es sich bei dem in dieser Arbeit durchgeführten Versuch nicht um ein Experiment, sondern um ein Quasi-Experiment (TU Dresden Juniorprofessur für Methoden der Psychologie & Kognitive Modellierung, 2017). Gründe dafür sind, dass es keine Kontrollgruppe gab und die

im Folgenden vorgestellten Störvariablen. Diese konnten nicht eliminiert werden, da es sich um ein Feldexperiment handelt. Um die Vorgehensweise der Plattformen gegen gefälschte Bewertungen als Laborexperiment zu testen, wäre eine tiefgehende Zusammenarbeit mit den Plattformen nötig gewesen.

3.8.1 Störvariablen

Da das Experiment als Feldexperiment durchgeführt wurde, ist prinzipiell mit Störvariablen, auch unbekanntes Störvariablen, zu rechnen. Grundsätzlich ist der Autor dieser Arbeit und der in ihrem Rahmen abgegebenen Bewertungen kein echter Spammer/Fälscher und hatte für die Abgabe von gefälschten Bewertungen andere Motivationen als Autoren echter gefälschter Bewertungen. Die im Experiment abgegebenen gefälschten Bewertungen könnten daher nicht authentisch sein. Auch haben die Einschränkungen zur Kontrolle der Folgen des Experiments für die Betroffenen (s. Kapitel 3.2) dazu geführt, dass teilweise andere Bewertungen abgegeben wurden als es echte Fälscher getan hätten. Ein entscheidender Unterschied zu realen gefälschten Bewertungen ist, dass nur je eine Bewertung pro Bewertungsgegenstand abgegeben wurde. Darüber hinaus wurden alle Bewertungen zu einem unterschiedlichen Zeitpunkt abgegeben. Dieser kann Einfluss auf die Freischaltung haben, etwa wenn bei einer Prüfung durch Menschen verschiedene Mitarbeiter zuständig sind oder sich im Zeitraum der Durchführung des Experiments Faktoren, die bei den Plattformen intern für den Umgang mit Bewertungen relevant sind, ändern.

Eine weitere Störvariable ist, dass bis auf wenige Ausnahmen bei den Blindtexten keine Bewertung mehrmals abgegeben wurde. So mussten die Plattformen zwar über gleich klassifizierte Bewertungen entscheiden, es handelte sich jedoch um unterschiedliche Texte, da diese zum jeweiligen Bewertungsgegenstand passen sollten. Es ist daher möglich, dass einige Bewertungen Faktoren enthielten, die neben dem beabsichtigten Faktor Einfluss auf den Online-Status hatten. Bei anderen, vermeintlich vergleichbaren Bewertungen kann dies nicht der Fall gewesen sein. Auch wurden rückwirkend betrachtet bei einem kleinen Teil der 192 abgegebenen Bewertungen die gewünschten Klassen nicht eindeutig getroffen.

Ebenfalls eine Störvariabel könnte es sein, dass Bewertungen zur Kontrolle ihres Online-Status immer wieder aufgerufen wurden. Dies könnte Aufmerksamkeit auf diese gelenkt haben. Um den Störvariablen entgegen zu wirken, wurde Randomisierung angewandt. So wurden die Bewertungen ohne Beachtung auf welcher Plattform sie veröffentlicht werden sollten verfasst und damit zufällig verteilt.

Bei der Messung konnten Störvariablen, anders als beim Verfassen der Bewertungen, vermieden werden: Der Onlinestatus einer Bewertung ist objektiv messbar und es besteht keine Gefahr der Fehlinterpretation.

3.8.2 Validität

Die externe Validität des Experiments ist hoch. Da das Experiment bereits in einer realen Umgebung stattgefunden hat, ist die Übertragbarkeit dorthin gegeben. Die interne Validität fällt dagegen wesentlich geringer aus. Da nur der Online-Status einer Bewertung gemessen werden konnte und nicht, wie die Plattformen diesen intern begründet haben, kann nicht ausgeschlossen werden, dass in einigen Fällen ein Störfaktor für den Online-Status entscheidend war.

3.8.3 Fazit zur Reflexion der Methodik

Um das Untersuchungsziel dieser Arbeit, den Umgang der Plattformen mit gefälschten Bewertungen, zu bearbeiten, ist ein Feldexperiment sinnvoll und ohne Kooperationen mit den Plattformen unumgänglich. Damit einher gehen Störvariablen, die nicht eliminiert werden können. Dies muss bei der Betrachtung der Ergebnisse beachtet werden. Trotz der niedrigeren internen Validität liefert das Experiment jedoch interessante und zur Beantwortung der Forschungsfrage zielführende Ergebnisse. Da eine hohe Anzahl an Bewertungen abgegeben wurde, ist davon auszugehen, dass es sich bei den damit gesammelten Erkenntnissen nicht nur um individuelle Beispiele handelt, sondern eine Übertragbarkeit auf die Gesamtmenge der bei den Plattformen verfassten gefälschten Bewertungen möglich ist.

4 Auswertung

Die in der Durchführungsphase des Experiments gewonnenen Rohdaten sind in einer Excel-Datei im Anhang dieser Arbeit zu finden. Im Folgenden werden diese zusammengefasst und interpretiert.

4.1 Auswertung des Experiments

Grundlegend ist festzuhalten, dass keine der verfassten Bewertungen im Verlauf der Messphase gelöscht wurde. Wenn eine Bewertung einmal online war (gemessen einen Tag nach dem Verfassen bzw. bei Zalando teilweise zwei Tage nach dem Verfassen) ist sie in der zweiwöchigen Messphase dauerhaft freigeschaltet gewesen. Bewertungen, die 14 Tage nach ihrem Verfassen nicht online waren, waren dies bereits bei der Ersten Messung ein/zwei Tage nach dem Verfassen nicht.

Im Folgenden werden die vier untersuchten Plattformen zunächst einzeln betrachtet:

4.1.1 Facebook

Unter den Facebook-Konten gibt es insofern einen Sonderfall, als dass ein Facebook-Konto der Klasse 1 (Wenig verdächtig), welches Bewertungen der Kategorie Betrug 2 abgegeben hatte, im Messzeitraum gesperrt wurde. Um das Konto wieder zu aktivieren, sollte die vermeintliche Inhaberin ein Foto von ihrem Gesicht hochladen. Als Grund für die Sperrung wurden verdächtige Login-Aktivitäten genannt. Das eigentliche Ziel des Kontos, wenig verdächtig zu erscheinen, wurde demnach verfehlt. Grund dafür dürfte sein, dass es innerhalb von zwei Tagen drei Logins von drei verschiedenen Geräten gab. Eine vermeintlich realistische Nutzung in diesen Sessions hat nicht gereicht, um Facebook von der Authentizität des Kontos zu überzeugen. Mit der Sperrung des Kontos waren auch die darüber verfassten Bewertungen nicht mehr öffentlich sichtbar, was in der folgenden Auswertung jedoch nicht als „Bewertung offline“ gewertet wurde. Grund dafür ist, dass der entscheidende dritte Login (nach Erstellung und Bewertungsabgabe), in dessen Session das Konto gesperrt wurde, bei anderen Konten nicht vorgenommen wurde. Es ist davon auszugehen, dass das Konto ohne diesen dritten Login nicht als verdächtig eingestuft worden wäre, wie es bei den anderen Konten der Fall war. Daher wurden die Bewertungen dieses Accounts aus der Grundgesamtheit der Bewertungen herausgerechnet (45 statt 48 Facebook-Bewertungen und 189 statt 192 Bewertungen insgesamt).

Keine der bei Facebook verfassten Bewertungen (außer wie erläutert die des gesperrten Kontos) wurde vollkommen gelöscht. Alle Bewertungen waren zwei Wochen nach dem Verfassen über ihren Direktlink abrufbar. Da Bewertungen jedoch in den seltensten Fällen über einen Direktlink aufgerufen werden, wurde zusätzlich die Sichtbarkeit einer Bewertung auf der bewerteten Seite

überprüft. 3 der 45 Facebook-Bewertungen waren nicht über die dazugehörigen Seiten aufrufbar. Diese wurden daher als offline gezählt. Facebook erreicht somit eine Löschquote von 6,7% der eingereichten gefälschten Bewertungen. Auffällig ist, dass die drei nicht sichtbaren Bewertungen sich auf sogenannte „Inoffizielle Seiten“ beziehen. Diese Facebook-Seiten haben keinen Administrator, sondern wurden beispielsweise aus Landkarten-Einträgen oder Wikipedia-Artikeln zu Orten automatisch generiert. Auf diesen Seiten zeigt Facebook nicht alle dafür abgegebenen Bewertungen an. Dies lässt sich leicht überprüfen, da die angezeigte Anzahl von Bewertungen für inoffizielle Seiten in vielen Fällen höher liegt als die tatsächlich angezeigte Anzahl an Bewertungen, wenn man diese händisch nachzählt. Facebook scheint demnach bei Seiten ohne Administrator weniger Inhalte an alle Nutzer auszuspielen als bei Seiten, denen ein Facebook-Konto zur Administration zugeordnet ist.

Da nur drei Bewertungen insgesamt von Facebook nicht ausgespielt wurden, sind keine Tendenzen, inwieweit dies mit dem Kontentyp oder der Klasse der Bewertung zusammenhängt, erkennbar. Nicht für alle Nutzer verfügbar gemacht wurden:

- eine Bewertung der Klasse **Spam 1** eines **wenig verdächtigen** Kontos mit dem Text „Lorem Ipsum...“,
- eine Bewertung der Klasse **Spam 2** eines **verdächtigen** Kontos, die nicht auf persönlichen Erfahrungen basierte („Habe gehört, dass...“) sowie
- eine Bewertung der Klasse **Betrug 2** eines **verdächtigen** Kontos, die einen inhaltlichen Fehler enthielt („Eintritt zu hoch!“ bei einer kostenlos zugänglichen Sehenswürdigkeit).

Da pro Bewertungsklasse bei Facebook zwölf Bewertungen (in der Klasse „Betrug 2“ neun Bewertungen) und so insgesamt 45 Bewertungen abgegeben wurden ist eindeutig, dass Facebook den größten Teil der gefälschten Bewertungen nicht als solche erkannt oder zumindest nicht gelöscht hat.

4.1.2 Google Maps

Bei Google Maps war der Online-Status einer Bewertung eindeutiger zu beurteilen als bei Facebook. Entweder war eine Bewertung über den Direktlink zum Autorenprofil und auf der bewerteten Seite einsehbar oder eine Bewertung war komplett gelöscht. Letzteres war bei zwei der 48 Bewertungen der Fall, was einer Löschquote von 4,2% entspricht. Die beiden gelöschten Bewertungen stammten von Konten der Klassen „Wenig verdächtig“ sowie „Verdächtig“ und enthielten beide Texte, die nach den Kriterien der Kategorie Spam 1 verfasst wurden. Eine Bewertung enthielt dabei sämtliche Zeichen des Alphabets („aäb cde fgh ijk lmn oöp qrsß tuü vwx yz AÄBC DEF GHI JKL MNO ÖPQ RST UÜV WXYZ“), die andere den Text

„ISARTOR\$EVERGEILSTERORTINMUNICH\$ITYYEAR!“. Google Maps hat damit zwar eine Spam 1-Bewertung mehr erkannt und gelöscht als Facebook, jedoch blieben 46 Bewertungen, davon 10 in der Klasse Spam 1 mit zum Teil trivialen Spam-Texten wie dem von Facebook erkannten „Lorem Impsum“, unerkant.

4.1.3 Google Play Store

Im Google Play Store wurden sechs von 48 Bewertungen nicht veröffentlicht (12,5%). Ein Zusammenhang mit dem Kontotyp ist auch hier nicht erkennbar, da Bewertungen von allen Kontentypen gelöscht wurden. Der Typ „Wenig verdächtig“ war mit vier gelöschten Bewertungen sogar stärker betroffen als die anderen zwei Typen mit jeweils einer Löschung.

Bei den Klassen der Bewertungen ist die Löschung (in geringem Ausmaß) auch klassenübergreifend erfolgt. So wurden folgende Bewertungen gelöscht:

- eine Bewertung eines **wenig verdächtigen** Kontos der Klasse **Betrug 1**, bei der als Freitext nur der Name eines Albums und der Plattenfirma kopiert wurde,
- eine Bewertung eines **wenig verdächtigen** Kontos der Klasse **Betrug 2**, in dem nicht das Produkt, sondern der Download bewertet wurde („... hat nicht geklappt“),
- zwei Bewertungen der Klasse **Spam 1** (Konto **wenig verdächtig**), die die Texte „####B#####a#####n#####n#####a...“ und „wollerocktdasausbeahmyeahr“ enthielten,
- eine Bewertung der Klasse **Spam 1** (Konto **sehr verdächtig**), welche wie bei Facebook zuvor alle Buchstaben des Alphabets hintereinander enthielt, sowie
- eine Bewertung der Klasse **Spam 2** eines **verdächtigen** Kontos, die auf keiner persönlichen Erfahrung beruhte („Konnte leider noch nicht reinhören...“)

4.1.4 Zalando

Zalando weist die im Experiment mit Abstand höchste Löschrquote von gefälschten Bewertungen auf. Dort wurden 21 der 48 abgeben Bewertungen nicht veröffentlicht. Die daraus resultierende Löschrquote von 43,8% ist somit mehr als zehnmahl so hoch wie die von Google Maps und fast viermahl so hoch wie die des Google Play Stores.

Folgende Bewertungen wurden nicht freigeschaltet:

- Konto wenig verdächtig, Betrug1 (Titel und Stoffzusammensetzung kopiert)
- Konto wenig verdächtig, Betrug1 („Der Versand erfolgte sehr schnell“)
- Konto wenig verdächtig, Spam1 (Alle Zeichen des Alphabets)
- Konto wenig verdächtig, Spam1 („Lorem Impsum...“)
- Konto wenig verdächtig, Spam1 („sehr gut qalitet“)
- Konto verdächtig, Betrug1 (Name und Preis kopiert)

- Konto verdächtig, Betrug1 (Name und Eigenschaften kopiert)
- Konto verdächtig, Betrug2 (Versandzeit bewertet)
- Konto verdächtig, Betrug2 (Versandzeit bewertet)
- Konto verdächtig, Betrug2 (Zahlung + Versand bewertet)
- Konto verdächtig, Spam1 (Kafka-Blindtext)
- Konto verdächtig, Spam1 („GUTERSCHNITTUNDTOLLEWASCHUNG SOMUSSEI...“)
- Konto verdächtig, Spam1 („Viktor jagte im komplett...“)
- Konto verdächtig, Spam2 („Gekauft, Bewertung folgt“)
- Konto sehr verdächtig, Betrug1 („Trendiger Look in hochwertiger Denim Qualität, perfekt für den Alltag oder die nächste Party!“)
- Konto sehr verdächtig, Betrug2 (Kartongröße bewertet)
- Konto sehr verdächtig, Betrug2 („Beschriftung nicht eindeutig“)
- Konto sehr verdächtig, Betrug2 (Verpackung bewertet)
- Konto sehr verdächtig, Spam1 (Blindtext Wikipedia-Artikel über Röcke)
- Konto sehr verdächtig, Spam1 („eone ser shcöne yeans paßd mir ser gud“)
- Konto sehr verdächtig, Spam1 (Goethe-Blindtext)

Bezogen auf die Kontotypen sind die Löschungen bei Zalando relativ gleichmäßig verteilt. Fünf Bewertungen der Klasse „Wenig verdächtig“ wurden gelöscht, neun der Klasse „Verdächtig“ und sieben der Klasse „Sehr verdächtig“. Dass die Löschraten von „Wenig verdächtig“ bis „Sehr verdächtig“ nicht ansteigend sind, deutet darauf hin, dass Zalando nur die Bewertungen an sich analysiert und nicht auf das dahinterstehende Konto bzw. das sonstige Nutzerverhalten eingeht.

Innerhalb der Bewertungsklassen variieren die Löschraten bei Zalando durchaus. Aufgrund der größeren Anzahl an gelöschten Bewertungen ist dies im Gegensatz zu den anderen Plattformen auch sinnvoll ausdifferenzierbar (siehe Abbildung 1). So wurden in der Kategorie Spam 1 (Blindtexte etc.) 75% der abgegebenen Bewertungen gelöscht, in der Kategorie Spam 2 (keine persönliche Erfahrung) jedoch nur 8,3%. Dazwischen lagen die Kategorien Betrug 1 (Produktinfos kopiert) mit 41,7% Löschraten und Betrug 2 (inhaltliche Fehler oder inhaltslose Bewertungen) mit einer Löschraten von 50%.

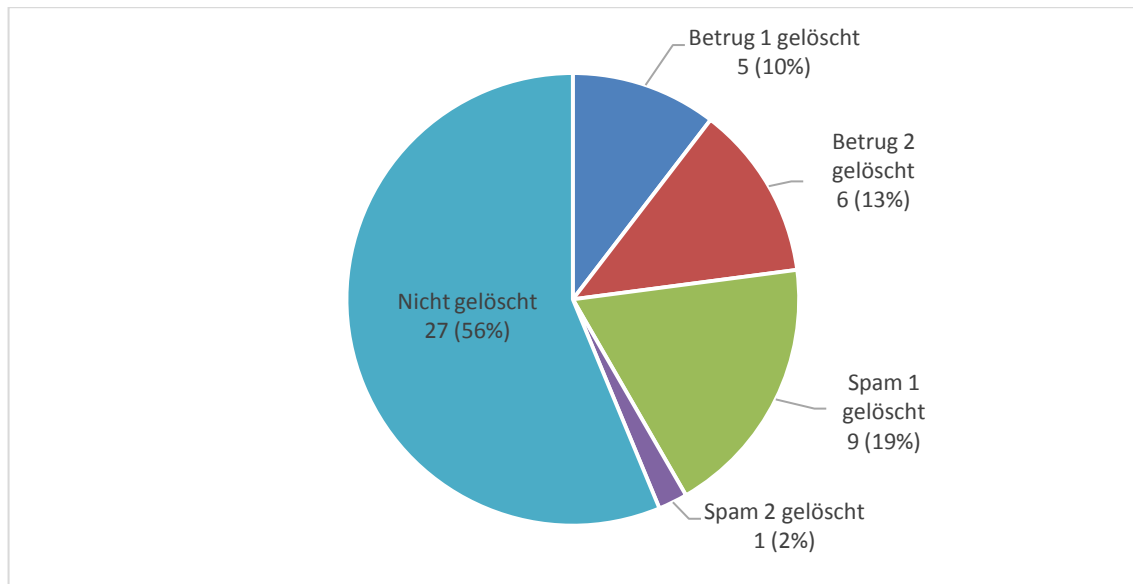


Abbildung 1: Anzahl und Anteil der ungelöschten und gelöschten Bewertungen bei Zalando

Ob über die Freischaltung einer Bewertung bei Zalando durch einen Algorithmus, durch Mitarbeiter oder durch beides entschieden wird, ist nur anhand der (nicht) gelöschten Bewertungen schwer einschätzbar. Auffällig ist, wie gut es Zalando schafft, Bewertungen auszusortieren, die keine Meinung zum Produkt enthalten. Konsequenterweise gelingt dies zwar nicht (beispielsweise wurden viele kopierte Produktbeschreibungen freigeschaltet), alle Bewertungen, die sich einzig auf Versand und Verpackung der Produkte bezogen, wurden jedoch ebenso wie fast alle Blindtexte gelöscht. Freigeschaltet wurden währenddessen alle Bewertungen, die explizit keine persönliche Meinung basierend auf eigener Erfahrung mit dem Produkt enthielten. Grund dafür könnte Versagen der Kontrolle von Zalando sein. Es ist aber auch möglich, dass Zalando beispielsweise Bewertungen, die ein Produkt nur anhand des Produktbildes, der Meinung Dritter oder dem vorhandenen Kaufwunsch beurteilen, hilfreich findet und den Kunden daher anzeigen möchte. Öffentliche Bewertungsrichtlinien, in denen dies erklärt sein könnte, gibt es bei Zalando wie in Kapitel 2.6 erläutert nicht. Darüber hinaus fällt auf, dass Zalando alle Bewertungen mit groben Rechtschreibfehlern (Spam 1 exklusive Blindtexte) gelöscht hat. Möglicherweise werden Bewertungen mit Wörterbuch-Listen abgeglichen oder (teilweise) von Menschen gelesen.

4.1.5 Plattformübergreifende Auswertung

Insgesamt wurden 32 der 189 Bewertungen von den Plattformen nicht freigeschaltet bzw. nicht für alle Nutzer zugänglich gemacht (siehe Abbildung 2). Das entspricht einer Löschquote von 16,9%, die jedoch durch Zalando im Vergleich zu den anderen drei Plattformen wesentlich höherer Löschquote (siehe Abbildung 3) mehr als verdoppelt wird. Facebook, Google Maps und Google Play Store erreichen zusammen lediglich eine Löschquote von 7,6%.

Die insgesamt am häufigsten gelöschte Bewertungsklasse ist Spam 1 mit einer Löschquote von 31,3% (15 von 48). Danach folgen Betrug 2 mit 17,8% (8 von 45) und Betrug 1 mit 12,5% (6 von 48). Bewertungen der Kategorie Spam2 wurden zu 6,3% gelöscht (3 von 48).

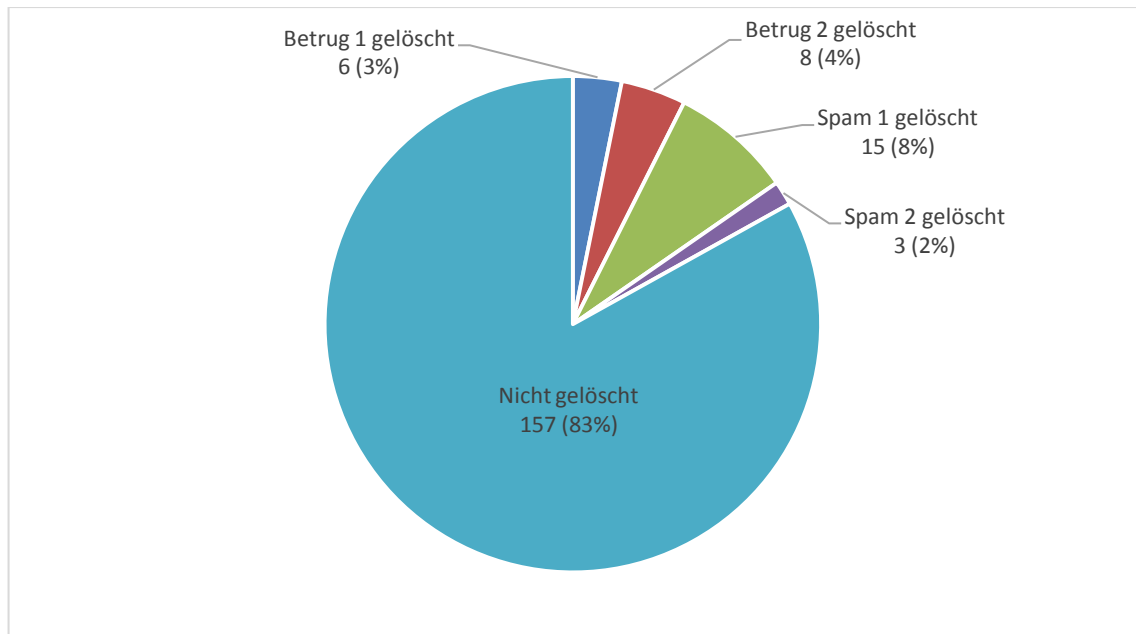


Abbildung 2: Anzahl der plattformübergreifend gelöschten und ungelöschten Bewertungen

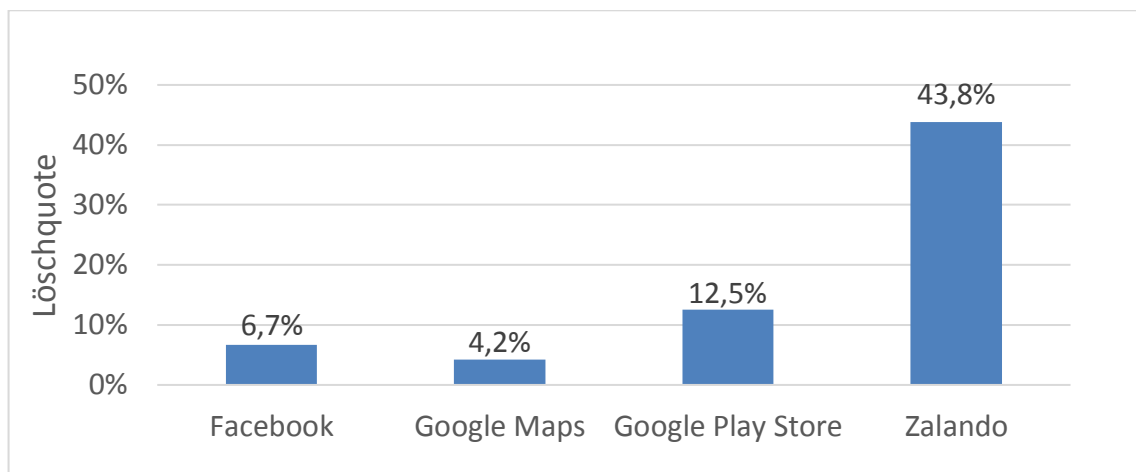


Abbildung 3: Löschquote der abgegebenen Bewertungen nach Plattformen

Auch plattformübergreifend ist keine Steigerung der Löschquote vom Kontentyp „Wenig verdächtig“ zum Kontentyp „Sehr verdächtig“ erkennbar (siehe Abbildung 4). Bei wenig verdächtigen Konten wurde in 35,5% (11 von 31) der Fälle eine oder mehrere Bewertungen gelöscht, bei verdächtigen Konten waren es 40,6% (13 von 32) und bei den als sehr verdächtig angelegten Konten 25% (8 von 32). Die niedrigste Löschquote weisen somit die Konten auf, bei denen die Erstellung von einer bis drei Bewertungen unmittelbar nach der Kontenerstellung die

einzigste Kontenaktivität war. Dies deutet darauf hin, dass die Plattformen sich bei der Analyse von Bewertungen nur wenig oder gar nicht nach dem dahinterstehenden Konto richten oder dass die Abstufungen der im Experiment angelegten Konten nicht eindeutig genug war.

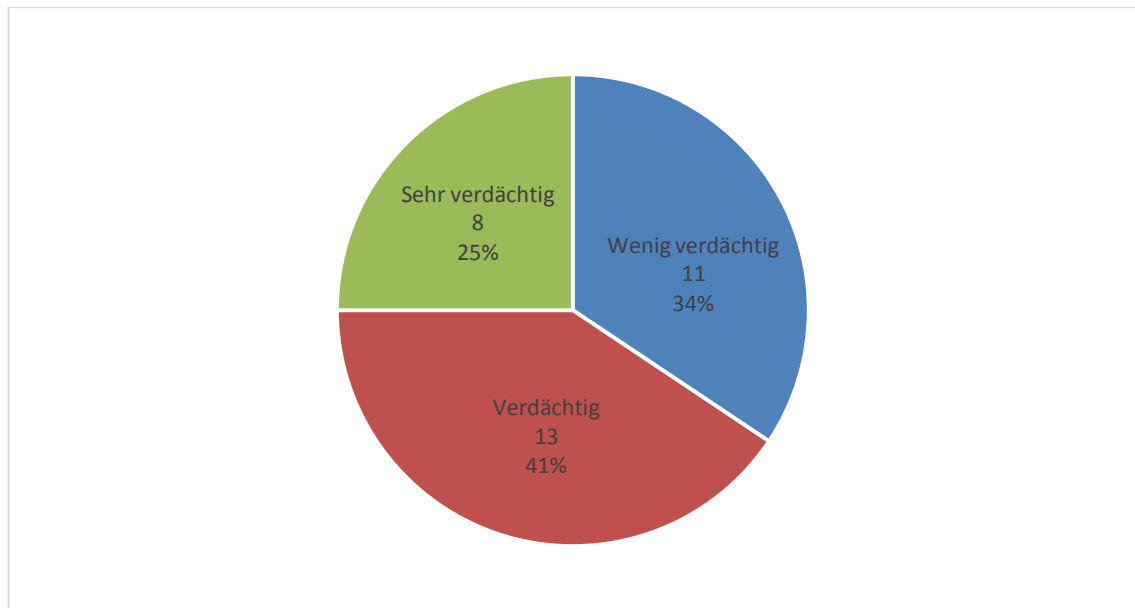


Abbildung 4: Plattformübergreifende Löschung aufgeschlüsselt nach Klassen der Konten

Neben den verschiedenen Kontotypen und Bewertungsklassen können die Konten auch nach der Anzahl der Bewertungen eingeteilt werden. Wie in Kapitel 3.6 dargestellt, wurde über die Hälfte der Konten jeweils eine Bewertung abgegeben und über die andere Hälfte der Konten jeweils drei Bewertungen. In der Auswertung konnte zwischen diesen zwei Kontengruppen jedoch kein nennenswerter Unterschied festgestellt werden. Zehnmal wurde eine Bewertung von einem Konto, über das eine einzige Bewertung abgegebene wurde, gelöscht und dreizehnmal wurde eine oder mehr Bewertungen von einem Konto mit drei Bewertungen gelöscht (von 96 Konten dieser Gruppe). Auch dies deutet darauf hin, dass eine einzelne Bewertung für die Plattformen eher im Vordergrund steht als das Verhalten des Kontos um sie herum.

Ebenfalls gleichmäßig verteilt wurde beim Abgeben der Bewertungen die Anzahl der vergebenen Sterne (Ausnahme Kategorie Betrug 1). Die Messergebnisse jeweils zwei Wochen nach Abgabe der Bewertungen zeigen, dass die Anzahl der vergebenen Sterne keinen Einfluss auf die Löschraten der Plattformen hat. Insgesamt wurden 15 Rezensionen mit negativer Bewertung (1 oder 2 Sterne) gelöscht und 17 Bewertungen mit positiver Bewertung (4 oder 5 Sterne). Die Löschraten sind also vergleichbar.

4.1.6 Sichtbarkeit der abgegebenen Bewertungen

In den vorherigen Abschnitten wurden die Messwerte, ob eine Bewertung für alle Nutzer einer Plattform sichtbar ist, dargelegt und interpretiert. In einigen Fällen muss dabei jedoch von einer „theoretischen Sichtbarkeit“ ausgegangen werden. Grund dafür sind die Bewertungen, die von anderen Nutzern vor und nach der Veröffentlichung einer gefälschten Rezension abgegeben wurden, sowie die Darstellungsweisen der Plattformen. Ein extremes Beispiel ist der Eintrag des Brandenburger Tors auf Google Maps. Dieses erhält wie andere international bekannte und viel besuchte Sehenswürdigkeiten täglich Dutzende Bewertungen. In der Sortierung „Neueste“ werden neu verfasste Bewertungen für das Brandenburger Tor sehr schnell soweit nach unten verdrängt, dass sie nach kurzer Zeit für normale Nutzer praktisch nicht mehr sichtbar sind. Wahrscheinlicher ist so, dass eine Bewertung über die Sortierung „Am hilfreichsten“ zu finden ist. Dort werden jedoch in erster Linie Bewertungen von erfahrenen Autoren („Local Guides“) angezeigt. Dass ein Nutzer auf Google Maps eine im Rahmen dieses Experiments abgegebene Bewertung zu sehen bekommt, ist bei häufig bewerteten Orten sehr gering. Lediglich in die Durchschnittsbewertung eines Ortes gehen die vergebenen Sterne marginal ein.

Bei den anderen Plattformen sowie weniger oft bewerteten Orten auf Google Maps sieht die Situation durchaus anders aus. So lassen sich bei Facebook neue Bewertungen in der Regel schnell über die Sortierung „Neueste“ finden. Dass deutschsprachige Facebook-Seiten innerhalb kurzer Zeit viele Bewertungen erhalten, scheint den Beobachtungen des Experiments nach eher ungewöhnlich zu sein und passiert nur bei bestimmten Anlässen wie Veranstaltungen.

Bei Zalando werden die Bewertungen unveränderbar in der Reihenfolge ihres Eingangs angezeigt (neueste zuerst). Da dort viele Artikel nur über wenige Bewertungen verfügen, ist die Sichtbarkeit von neuen gefälschten Bewertungen sehr gut und lange gegeben. Im Google Play Store kann die Sortierung der Bewertungen (außer bei Apps) durch den Nutzer ebenfalls nicht verändert werden. Sortiert wird dort nach einem auf verschiedenen Faktoren basierenden Algorithmus und nicht (allein) nach dem Datum einer Bewertung.

Die Sichtbarkeit der im Experiment bei Google Play abgegebenen Bewertungen war sehr unterschiedlich. In vielen Fällen wurden die Bewertungen sehr weit unten in der Liste der Bewertungen aufgeführt, in einigen Fällen war jedoch auch das komplette Gegenteil der Fall und eine Bewertung wurde unter den ersten Rezensionen ganz oben angezeigt. Bei diesen Bewertungen handelte es sich um längere Blindtexte ohne Bezug zum Bewertungsgegenstand. Google scheint im Play Store die Länge einer Bewertung als Kriterium für ihre Qualität mit einzubeziehen. In einigen dieser Fälle kam es zudem dazu, dass die Blindtext-Bewertungen von

Benutzern als „Hilfreich“ (Daumen nach oben) bewertet wurden, was ihre Position in der Sortierung verstärkt oder gesichert haben dürfte (siehe Abbildung 5).

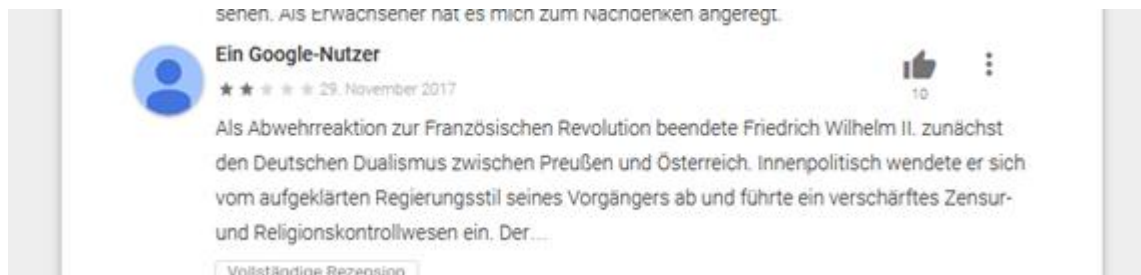


Abbildung 5: Ein Blindtext als Bewertung eines Kinderfilms im Google Play Store wurde zehnmals als hilfreich bewertet

4.2 Folgeexperiment: Melden der Bewertungen

Nach Abschluss der Messung für das Experiment wurden die Bewertungen, sofern es möglich war, an die Plattformbetreiber gemeldet. Ziel war es herauszufinden, ob die Plattformen eine Bewertung, die zunächst nicht gelöscht wurde, nach der Meldung löschen. Bei Zalando ist es nicht vorgesehen, dass Nutzer Bewertungen melden können. Daher wird die Plattform in diesem Folgeexperiment nicht mehr betrachtet.

Für das Melden einer Bewertung bei Google Maps ist kein Google-Account nötig, lediglich eine E-Mail-Adresse muss hinterlegt werden (siehe Abbildung 6). Dennoch wurden die Meldungen in verschiedenen Browserkonfigurationen des Hauptexperiments durchgeführt, die Konten haben also gegenseitig ihre Bewertungen gemeldet. Die Rezensionen wurden bei der Meldung als „Werbung oder Spam“ oder als „weicht vom Thema ab“ klassifiziert.



Richtlinienverstoß melden

Wir bei Google nehmen jegliche Art von Missbrauch unserer Dienste sehr ernst. Geben Sie in den nachfolgenden Feldern an, warum der Kommentar Ihrer Ansicht nach gegen unsere Nutzungsbedingungen verstößt. Wir werden den Fall untersuchen und gegebenenfalls entsprechende Maßnahmen einleiten. Wir melden uns nur bei Ihnen, falls sich Rückfragen ergeben oder wir zusätzliche Informationen für Sie haben.

- Wenn Sie einen oder mehrere Erfahrungsberichte gefunden haben, deren Entfernung Ihrer Ansicht nach gemäß anwendbarem Recht gerechtfertigt ist, [folgen Sie der Anleitung auf dieser Seite](#), um eine rechtsrelevante Anfrage einzureichen.

E-Mail-Adresse: *

Art des Verstoßes: *

- Dieser Beitrag enthält Inhalte, die zu Hass aufrufen, gewalttätig oder unangemessen sind.
- Dieser Beitrag enthält Werbung oder Spam.
- Dieser Beitrag weicht vom Thema ab.
- Dieser Beitrag enthält Interessenkonflikte.

Abbildung 6: Einen Richtlinienverstoß bei Google Maps melden

Bei Google Play ist für das Melden von Bewertungen ein Login erforderlich. Auch hier wurden die Konten aus dem Hauptexperiment genutzt und jeweils ein Konto zum Melden von etwa fünf Bewertungen verwendet. Die Meldung an sich unterscheidet sich je nach Medienart. Musik und Bücher werden gemeldet, indem der Nutzer den Mauszeiger über die Bewertung bewegen muss und auf ein Flaggen-Piktogramm klicken muss. Als Feedback wird die gemeldete Bewertung für den Nutzer anschließend ausgeblendet und es bleibt eine weiße Fläche. Bei anderen Medien im Google Play Store erfolgt die Meldung einer Rezension als Spam über ein Kontextmenü, welches sich für jede Bewertung aufklappen lässt.

Bei Facebook ist das korrekte Melden einer Bewertung der Klasse Spam 2 (beispielsweise „Ich kenne den Ort nicht, aber...“) nicht möglich. Im Dialog zum Melden einer Bewertung können nur „Es hat nichts mit dieser Seite zu tun“ oder verschiedene Aspekte der Gemeinschaftsstandards wie Gewalt, Bedrohung oder Drogenhandel als Grund zur Meldung ausgewählt werden. Alle im Experiment abgegeben gefälschten Bewertungen wurden daher als „Es hat nichts mit dieser Seite zu tun“ gemeldet, da dies unter den verfügbaren Optionen am ehesten zutraf. Es stellt sich jedoch die Frage, ob Facebook überhaupt ein Interesse daran hat, nicht hilfreiche gefälschte Bewertungen zu entfernen, solange sie keine in den Gemeinschaftsstandards genannten Aspekte verletzen.

14 Tage nach Meldung der Rezensionen wurde kontrolliert, ob diese noch online für alle Nutzer einsehbar sind. Bei Google Maps und im Google Play Store war dies durchgehend der Fall. Keine der Meldungen hat hier die Löschung einer Bewertung zur Folge. Möglicherweise wird Google erst bei mehreren Meldungen einer Bewertung aktiv oder die Überprüfung einer Bewertung

dauerte zum Versuchszeitpunkt länger als zwei Wochen. Es ist nicht davon auszugehen, dass Google sich bei allen Rezensionen bewusst dazu entschieden hat, diese nicht zu löschen. Da im Hauptexperiment kaum Bewertungen gelöscht wurden, befanden sich noch viele sehr offensichtliche Spam-Texte unter den gemeldeten Bewertungen.

Bei Facebook wurden 21 der 44 gemeldeten Bewertungen gelöscht (47,7%). Diese Bewertungen wurden somit nicht mehr auf den entsprechenden Seiten angezeigt und ihr Direktlink wurde zu einer Fehlerseite. Darüber hinaus wurde der Eigentümer eines Kontos in einem Popup an die Regeln für Bewertungen auf Facebook hingewiesen, welches bestätigt werden musste, um Facebook weiter nutzen zu können (siehe Abbildung 7).

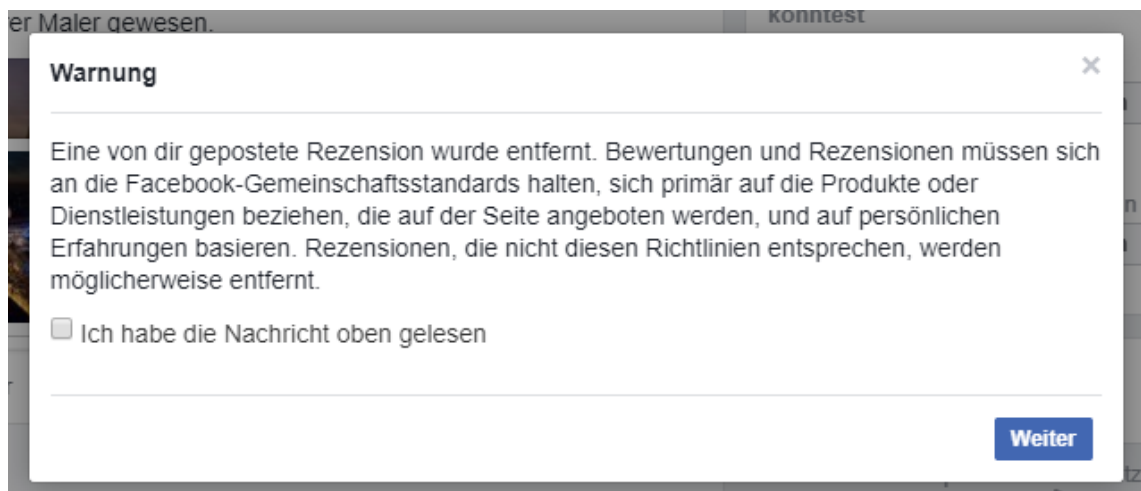


Abbildung 7: Facebook ermahnt den Autor einer gefälschten Bewertung

Die meisten Facebook-Bewertungen wurden innerhalb der Klasse Spam 2 gelöscht (8 von 12 bzw. 75%). Gemeldete Spam 1-Bewertungen wurden in 50% der Fälle entfernt (6 von 12), Betrug 1-Bewertungen zu 27,3% (3 von 11) und Betrug 2-Bewertungen zu 22,2% (2 von 9). Bezogenen auf die verschiedenen Klassen von Konten wurden 8 von 14 Bewertungen der Konten des Typs „Sehr verdächtig“ (57,1%) sowie jeweils fünf Bewertungen der Klassen „Wenig verdächtig“ und „Verdächtig“ (41,7% bzw. 31,3%) gelöscht. Dies ist zwar eine andere Verteilung als beim Hauptexperiment, die Ausprägungen sind jedoch nicht stark genug um festhalten zu können, dass die Konten hinter einer Bewertung, anders als bei der Veröffentlichung, bei Facebooks Prüfung nach der Meldung zur Löschentscheidung herangezogen wurden.

Eine im Hauptexperiment abgegebene Facebook-Bewertung (Betrug 1/Wenig verdächtig) muss zudem in der kurzen Zeit zwischen der letzten Kontrolle des Hauptexperiments und dem Beginn des Folgeexperiments gelöscht worden sein. Als sie gemeldet werden sollte, war sie bereits nicht mehr online. Möglicherweise wurde sie von einem Nutzer gemeldet und daraufhin gelöscht. Allgemein scheint es jedoch eher selten vorzukommen, dass Nutzer gefälschte Bewertungen

zeitnah nach deren Veröffentlichung den Plattformen melden. Immerhin hätten auf diesem Weg im Messzeitraum des Hauptexperiments Bewertungen von Facebook gelöscht werden können, dies ist jedoch in keinem Fall geschehen.

4.3 Googles Registrierungsprozess

Wie in Kapitel 3.4 dargelegt, verwendet Google in seinem Registrierungsprozess einen Algorithmus, der individuell entscheidet, ob eine Verifikation per Telefonnummer zur Eröffnung eines Kontos notwendig ist. Aufgrund der vielen im Rahmen des Experiments angelegten Google-Konten konnten Rückschlüsse auf die Faktoren gezogen werden, die dazu führen, dass eine Verifikation verlangt wird.

Grundlegend ist zunächst die IP-Adresse bzw. der Anschluss, von dem ein Konto angelegt wird. Bei der Nutzung öffentlicher Hotspots und dem halböffentlichen DFN (Deutsches Forschungsnetz) wurde im Experiment stets eine Telefonnummernverifizierung gefordert. Bei der Nutzung von Privatanschlüssen und Mobilfunk war dies meistens nicht der Fall. Jedoch konnten hier trotz wechselnder IP-Adressen maximal etwa fünf Konten pro Tag angelegt werden. Möglicherweise gleicht Google IP-Adressräume und eventuell zusätzlich Gerätemerkmale mit vergangenen Registrierungen ab. Wenn innerhalb mehrerer Stunden zu viele Registrierungen in einem IP-Adressraum durchgeführt werden bzw. dies zusätzlich vom mutmaßlich gleichen Gerät aus geschieht, wird die zusätzliche Verifizierung verlangt.

Ursprünglich sollten Konten im Experiment mit deutschen Festnetznummern des VOIP-Anbieters Sipgate verifiziert werden. In der Durchführung hat sich jedoch herausgestellt, dass diese von Google zur Verifizierung nicht angenommen wurden. Möglicherweise akzeptiert Google grundsätzlich keine VOIP-Nummern, die sich kostenlos online einrichten lassen. Nach Aussage des Sipgate Kundenservice könnte Google bei der Bundesnetzagentur eingesehen haben, welche Nummern zu welchem Anbieter gehören. Ob Google dies tut, ist Sipgate aber nicht bekannt.

5 Zusammenfassung

Bevor ein Fazit zum Ergebnis der Arbeit gezogen wird, werden im Folgenden die Ergebnisse des Experiments zusammengefasst dargestellt.

5.1 Erstellung der Konten

Die Konten ließen sich im Experiment relativ leicht erstellen. Bei Zalando können Bewertungen gar ohne irgendeine Form der Nutzer-Verifizierung nach Eingabe einer fiktiven E-Mail-Adresse verfasst werden, was den möglichen Missbrauch stark vereinfacht. Bei Facebook muss zumindest eine E-Mail-Adresse (oder Telefonnummer) bestätigt werden. Liegt diese vor, ist das Erstellen eines voll funktionsfähigen Kontos jedoch kein Problem. Erst wenn mehrere verdächtige Aktivitäten wie im Experiment bei einem Konto die hohe Zahl an verschiedenen Geräten und IP-Adressen zusammenkommen, verlangt Facebook eine komplexere Verifizierung. Die Erstellung von Google-Konten war im Experiment am schwierigsten, da dort teilweise eine Verifizierung per SMS-Code zum Abschließen des Registrierungsvorgangs verlangt wird. Dennoch konnten knapp 50 Konten in weniger als zwei Wochen erstellt werden. Besonders wenn Konten über einen längeren Zeitraum hinweg erstellt werden, stellen weder Google noch die anderen Plattformen einem Bewertungsfälscher bei der Kontenerstellung ernsthafte Hürden in den Weg.

5.2 Verfassen von Bewertungen

Wie im vorherigen Kapitel erläutert, ist die Löschquote von gefälschten Bewertungen der Plattformen sehr gering. Besonders bei Facebook und Google Maps sind mehr als 90% der Spam oder Betrug simulierenden Bewertungen veröffentlicht wurden. Dies zeigt, dass es sehr leicht ist, auf den Plattformen gefälschte Bewertungen zu verfassen. Bei Google Play war die Löschquote etwas höher, bei Zalando deutlich höher. Dort wurden jedoch größtenteils Bewertungen ohne Produktbezug wie Blindtexte und Bewertungen des Versands gelöscht. Bewertungen, die sich auf das Produkt bezogen, jedoch offensichtlich auf keiner persönlichen Erfahrung beruhten, wurden bei Zalando ebenfalls zu über 90% freigeschaltet. Dies war auch der Fall, wenn eine solche Bewertung von einem neuen Konto ohne andere Aktivität abgegeben wurde. Es ist daher davon auszugehen, dass es nicht komplizierter wäre, das Meinungsbild zu einem Produkt, welches sich durch die Kundenbewertung ergibt, zu verzerren. Da viele Produkte auf Zalando nur über wenige Rezensionen verfügen und zudem immer die neuesten Bewertungen zuerst angezeigt werden, ist der Aufwand durch die niedrige Anzahl der dazu notwendigen Konten zumindest für eine kurz- und mittelfristige Manipulation überschaubar.

Anders sieht es in vielen Fällen bei Google Maps aus. Es scheint, als wolle Google gefälschten oder nicht hilfreichen Bewertungen entgegenwirken, indem insgesamt eine möglichst große Anzahl an Bewertungen angeboten wird. Um diese zu generieren fordert die Google Maps App Nutzer zu Rezensionen auf und regelmäßige Rezensenten werden etwa durch die „Local Guide“-Kennzeichnung belohnt (Amadeo, 2015). Bei beliebten Orten geht dieses Konzept auf: Nicht hilfreiche Bewertungen gehen im Rauschen der vielen abgegebenen Bewertungen unter und haben auch auf die Gesamtbewertung nur verschwindend geringen Einfluss. Der Aufwand zur Verzerrung der Gesamtbewertung eines von echten Nutzern oft besuchten Ortes wäre aufgrund der Vielzahl an dafür nötigen Konten sehr hoch. Bei weniger oft besuchten Orten ist dies jedoch nicht der Fall und Googles Löschquote gefälschter Bewertungen bei Maps war im Experiment mit 4,2% die niedrigste der untersuchten Plattformen. Dazu kommt, dass Bewertungen auf Google Maps aufgrund der großen Zahl der Nutzer der Google-Websuche sowie von Google Maps selbst von sehr vielen Nutzern wahrgenommen werden können. Das Fälschen von Google Maps Bewertungen scheint nach den Ergebnissen des Experiments nicht schwer und aufgrund der hohen Sichtbarkeit von Bewertungen beispielsweise für Inhaber von verhältnismäßig wenig besuchten Orten durchaus reizvoll.

Facebook hat seine Löschquote im Nachexperiment durch die Löschung nach Meldung deutlich erhöht. Dies lässt auf die dort angewendete Strategie schließen: Inhalte werden zunächst ungeprüft oder nur schwach algorithmisch geprüft freigeschaltet. Erst wenn diese gemeldet wurden, werden sie intensiver überprüft und ggf. gelöscht. Dennoch waren zwei Wochen nach Meldung der Bewertungen und über einen Monat nach dem Verfassen dieser immer noch etwas mehr als die Hälfte der Bewertungen online. Auch bei Facebook kann somit nicht davon gesprochen werden, dass gefälschte Bewertungen konsequent aussortiert werden.

Google Play war in verschiedener Hinsicht die heterogenste Plattform im Experiment. Zum einen war die Sichtbarkeit der Bewertungen hier sehr unterschiedlich. Mal wurden sie in der Sortierung weit nach hinten geschoben, mal weit nach vorne. In 12,5% der Fälle wurden Bewertungen gelöscht. Zum anderen war es schwierig, innerhalb der Gruppe der gelöschten Bewertungen Muster zu erkennen. Die Klasse Spam 1 wurde, wie es auch plattformübergreifend der Fall war, häufiger gelöscht. Darüber hinaus waren aber alle Klassen ähnlich stark betroffen und Bewertungen weniger verdächtiger Konten wurden nicht öfter gelöscht als die sehr verdächtiger Konten. Insgesamt kann jedoch festgehalten werden, dass bei 87,5% nicht gelöschten Bewertungen nach vier Wochen inkl. Meldung der Bewertungen dazwischen das Verfassen von gefälschten Bewertungen im Google Play Store nicht schwierig ist.

5.3 Methodik

Mit Hilfe der gewählten Methodik konnte herausgefunden werden, dass die untersuchten Plattformen anfällig für gefälschte Bewertungen sind. Die Forschungsfrage der Arbeit konnte so beantwortet werden. Aufgrund der hohen Anzahl von abgegebenen Bewertungen ist davon auszugehen, dass sich die Ergebnisse des Experiments verallgemeinern lassen. Dies trifft jedoch nur auf die untersuchten Typen von gefälschten Bewertungen zu. Wie gut die Plattformen beispielsweise Bewertungen aussortieren können, die authentisch wie eine tatsächliche Nutzermeinung verfasst wurden oder illegale Inhalte enthalten, wurde nicht untersucht. Auch das Verfassen von vielen gefälschten Bewertungen für einen Bewertungsgegenstand wurde nicht simuliert.

Die im Experiment vorhandenen Störvariablen sind bei der gewählten Methodik größtenteils unumgänglich. Bei einem Folgeexperiment könnte jedoch genauer darauf geachtet werden, dass die abgegebenen Bewertungen eindeutig den zugewiesenen Klassen entsprechen. Dies könnte beispielsweise erreicht werden, indem mehrere Personen an der Erstellung der Bewertungen beteiligt sind und diese gegenseitig kontrollieren oder fremde Bewertungen zur Kontrolle selbst klassifizieren.

6 Fazit

Das im Rahmen dieser Arbeit durchgeführte Experiment hat gezeigt, dass die Plattformen sehr anfällig für die im Experiment abgegebenen gefälschten Bewertungen sind. Weder gibt es bei der Erstellung von Konten nennenswerte Hindernisse, noch werden Bewertungen von den Plattformen so effektiv kontrolliert, dass das Erscheinen vieler gefälschter Bewertungen verhindert wird. Am ehesten schaffen es die Plattformen gegen offensichtliche Spam-Texte ohne Bezug zum Bewertungsgegenstand vorzugehen.

Wenn böswillige oder unehrliche Nutzer auf einer der untersuchten Plattform gefälschte Bewertungen abgeben wollen, können sie dies nahezu ungehindert tun. Besonders wenn die Abgabe von gefälschten Bewertungen über einen längeren Zeitraum hinweg erfolgt, ist nicht davon auszugehen, dass die Plattformen die gefälschten Bewertungen erkennen. Während es bei Zalando und teilweise im Google Play Store (dort zumindest bei den vielen von großen Konzernen eingestellten Artikeln im Film- und Musikbereich) eher unwahrscheinlich ist, dass im größeren Stil manipulierte Bewertungen verfasst werden, scheint dies bei Facebook und Google Maps plausibler. Dort ist die Anzahl der Personen, die z. B. als Eigentümer eines Ortes von gefälschten Bewertungen profitieren könnten, höher. So ist eher davon auszugehen, dass einige von ihnen Manipulationen am Bewertungssystem vornehmen.

Dass bei Zalando die meisten gefälschten Bewertungen aussortiert wurden, könnte daran liegen, dass hier Bewertungen relativ direkt mit dem eigenen Umsatz verknüpft sind. So werden ausschließlich Artikel bewertet, die der Nutzer über einen Button in direkter Nähe zu den Bewertungen erwerben kann. Die Bewertungen haben so eine beratende Funktion bezogen auf die Kaufentscheidung. Facebook und Google Maps bieten dagegen eher ein allgemeines Informationsportal zu unzähligen Orten/Seiten mit vielen weiteren Informationen an. Der Google Play Store befindet sich in einer ähnlichen Situation wie Zalando. Dafür spricht, dass hier fast dreimal so viele Bewertungen gelöscht wurden wie bei Google Maps vom selben Anbieter. Grundsätzlich dürften zwar alle Anbieter ein Interesse an authentischen Bewertungen haben, der direktere Zusammenhang zum Umsatz kann für die Plattformen jedoch ein zusätzlicher Anreiz für die Kontrolle der Bewertungen sein. Ein häufigeres Löschen von negativen Bewertungen konnte im Google Play Store und bei Zalando allerdings nicht festgestellt werden. Die Plattformen lassen sich negative Meinungen zu den von ihnen angebotenen Produkten also gefallen.

Bezogen auf die Auswahl der Plattformen im Experiment ist anzumerken, dass Bewertungen für alle Plattformen nicht Teil ihres Kerngeschäfts sind. Keine der Plattformen ist eine reine Bewertungsplattform, die von ihren Nutzern in erster Linie aufgrund der dort zu findenden

Bewertungen aufgerufen wird. Es ist möglich, dass reine Bewertungsportale, eingehende Bewertungen strenger prüfen, als es die Plattformen im Experiment getan haben. Dennoch sind die untersuchten Plattformen aufgrund ihrer hohen Anzahl an Nutzern wichtige Distributoren von Bewertungen für Orte und die dort angebotenen Produkte.

Diese vielen Nutzer dürften jedoch auch eine zentrale Schwierigkeit sein, die verhindert, dass Bewertungen effektiver auf Fälschungen untersucht werden. Je mehr Bewertungen bei einer Plattform eingehen, desto aufwendiger ist es, die Qualität dieser zu gewährleisten. Dennoch ist es erstaunlich, dass die Plattformen bei neu angelegten Konten nicht strenger sind. Ein Konto, das einzig dafür angelegt wurde eine Bewertung zu verfassen, dürfte nur in wenigen Fällen von einem echten Nutzer angelegt worden sein. Den untersuchten Plattformen scheint es demnach wichtiger zu sein, neuen Nutzern so wenig Hürden wie möglich in den Weg zu stellen und die Qualität von Bewertungen durch Quantität zu erreichen.

7 Ausblick

Für die Untersuchung des in dieser Arbeit behandelten Themas sind eine Reihe von anderen Experimenten denkbar und spannend. Wie bereits erwähnt, wäre es interessant zu untersuchen, wie reine Bewertungsplattformen mit gefälschten Bewertungen umgehen. So könnte die Hypothese, dass dort strenger kontrolliert wird, überprüft werden. Generell wäre die Untersuchung weiterer Plattformen interessant, da die ausgewählten vier Plattformen zwar sehr bedeutend sind, letztendlich aber nur einen kleinen Teil der Plattformen mit der Möglichkeit zum Verfassen von Bewertungen im Internet darstellen.

Jenseits der Untersuchung anderer Plattformen sind auch Veränderungen im Ablauf des im Rahmen dieser Arbeit durchgeführten Experiments denkbar. So könnten mehr Kriterien für das Verfassen von Bewertungen herangezogen werden. Beispielsweise linguistische Merkmale, welche in diesem Experiment nicht betrachtet wurden. Es könnte untersucht werden, ob Plattformen Bewertungen mit deutlichen Übertreibungen (z. B. sehr viele positive Adjektive) löschen. Allgemein könnten Kriterien entwickelt werden, die sich stärker auf die wissenschaftlichen Veröffentlichungen zur Erkennung gefälschter Bewertungen beziehen und die darin genannten Kriterien zur Erkennung bewusst berücksichtigen. Dass bereits im eher auf die Regeln der Plattformen fokussierten Experiment dieser Arbeit sehr viele Bewertungen nicht erkannt wurden, könnte dafür sprechen, dass dies bei potenziell komplizierteren Kriterien aus wissenschaftlichen Arbeiten noch stärker der Fall sein wird. Dies ist jedoch eine Vermutung, die es noch zu belegen gilt.

Ein interessantes Ergebnis des Experiments ist es, dass mehr Bewertungen von als „Wenig verdächtig“ angelegten Konten als von als „Sehr verdächtig“ angelegten Konten gelöscht wurden. Es wäre spannend herauszufinden, ob dies daran lag, dass die im Experiment vorgenommene Abstufung nicht eindeutig genug war oder ob die Plattformen das Konto hinter einer Bewertung nicht zur Analyse dieser betrachten. In einem entsprechenden Experiment könnten die Abstufungen zwischen den Konten größer sein, indem ältere Konten mit mehr Interaktion erstellt werden. Dies wäre mit der Verwendung vieler Mobilgeräte im Experiment möglich. Darüber hinaus könnten wesentlich ähnlichere Bewertungen von den Konten abgegeben werden, um Störvariablen durch verschiedene Texte zu vermindern und die Untersuchung stärker auf die Kontentypen zu fokussieren.

Allgemein lässt sich mit Blick in die Zukunft festhalten, dass das Problem gefälschter Bewertungen in absehbarer Zeit nicht verschwinden wird. Zwar gibt es immer wieder Fortschritte in der algorithmischen Erkennung gefälschter Bewertungen, eine vollständige Erkennung bieten diese Ansätze jedoch nicht. Zudem werden neueste Erkenntnisse der

Wissenschaft nicht zwangsläufig von den Plattformen integriert. Eine stärkere Authentifizierung des Autors einer Bewertung, beispielsweise mit einem elektronischen Personalausweis, könnte die Anzahl der gefälschten Bewertungen wahrscheinlich deutlich senken. Ob dies gesellschaftlich gewollt ist, ist jedoch eine andere Frage. Auf absehbare Zeit sind die Plattformen so weiterhin auf die Analyse eingehender Bewertungen angewiesen. Einen ausreichend hohen Einsatz zeigen – zumindest die untersuchten Plattformen – dabei derzeit aber nicht.

8 Literaturverzeichnis

Amadeo, Ron (2015) New Google Maps notification asks for your restaurant food photos, Verfügbar unter: <https://arstechnica.com/gadgets/2015/08/new-google-maps-notification-asks-for-your-restaurant-food-photos/> [abgerufen am: 02.01.2018].

Banerjee, Snehasish, Chua, Alton Y. K., (2014) Understanding the Process of Writing Fake Online Reviews. In: IEEE, Hrsg. Ninth International Conference on Digital Information Management (ICDIM), Phitsanulok, Thailand: IEEE S. 68-73.

Conrady, Roland (2014) Customer Reviews: kaufentscheidend, glaubwürdig, strategierelevant?, Verfügbar unter: <https://v-i-r.de/wp-content/uploads/2017/08/Studie-Hotelbewertungen-FHWorms.pdf> [abgerufen am: 12.01.2018].

EHI Retail Institute (2017) Top 100 umsatzstärkste Onlineshops in Deutschland, Verfügbar unter: <https://www.ehi.org/de/top-100-umsatzstaerkste-onlineshops-in-deutschland/> [abgerufen am: 05.12.2017].

Facebook (2017a) Company Info, Verfügbar unter: <https://newsroom.fb.com/company-info/> [abgerufen am: 05.12.2017].

Facebook (2017b) Gemeinschaftsstandards, Verfügbar unter: <https://www.facebook.com/communitystandards> [abgerufen am: 05.12.2017].

Facebook (2017c) Informationen zu Bewertungen deines Shops, Verfügbar unter: <https://www.facebook.com/business/help/1131057156934245> [abgerufen am: 05.12.2017].

Facebook (2017d) Wie setzt sich die Sternbewertung einer Seite zusammen?, Verfügbar unter: <https://www.facebook.com/help/500762053364226> [abgerufen am: 05.12.2017].

Facebook (2017e) Wie verfasse ich eine Bewertung oder Rezension für ein Unternehmen?, Verfügbar unter: <https://www.facebook.com/help/271128916342286> [abgerufen am: 05.12.2017].

Google My Business-Hilfe (2017) Richtlinien für Rezensionen, Verfügbar unter: <https://support.google.com/business/answer/2622994?hl=de> [abgerufen am: 07.12.2017].

Google Play (2017) Richtlinien zum Posten von Kommentaren Verfügbar unter: <https://play.google.com/about/comment-posting-policy.html> [abgerufen am: 07.12.2017].

Kuhr, Daniela (2013) Amazons zweifelhafte Sternchen, Verfügbar unter: <http://www.sueddeutsche.de/geld/exklusive-kundenbewertungen-amazons-zweifelhafte-sterne-1.1791159> [abgerufen am: 20.12.2017].

Lappas, Theodoros, (2012) Fake Reviews: The Malicious Perspective. In: Bouma et al., Hrsg. Natural Language Processing and Information Systems, Groningen, Niederlande: Springer S. 23-34.

Lim, E.-P. et. al., (2010) Detecting Product Review Spammers using Rating Behaviors. In: ACM, Hrsg. CIKM'10, Toronto: ACM

Liu, Bing (2010) Sentiment Analysis and Subjectivity In: Indurkha, N., Damerau F. J. , Handbook of Natural Language Processing, S. Verfügbar unter: <https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf> (abgerufen am: 16.01.2018).

Mowery, Keaton, Shacham, Hovav (2012) Pixel Perfect: Fingerprinting Canvas in HTML5, Verfügbar unter: <http://cseweb.ucsd.edu/~hovav/dist/canvas.pdf> [abgerufen am: 09.01.2017].

Mühlenbeck, Frank, Skibicki, Klemens (2010) Authentizität von Hotelbewertungsplattformen – wie mächtig und wie glaubwürdig ist User Generated Content?, In: Amersdorffer et al., Social Web im Tourismus, Berlin, Heidelberg: Springer Verlag, S. 62.

Mukherjee, A. et al. (2013) What Yelp Fake Review Filter Might Be Doing?, Verfügbar unter: <https://www.cs.uic.edu/~liub/publications/ICWSM-2013-Arjun-Spam.pdf> [abgerufen am: 16.01.2018].

Mukherjee, Arjun, (2015) Detecting Deceptive Opinion Spam using Linguistics, Behavioral and Statistical Modeling. In: Linguistics, Association for Computational, Hrsg. 53rd Annual Meeting of the ACL and the 7th IJCNLP 26-31 Juli 2015, Beijing, China: Association for Computational Linguistics S. 21-22.

Mukherjee, Arjun, Liu, Bing, Glance, Natalie, (2012) Spotting Fake Reviewer Groups in Consumer Reviews. In: Committee, International World Wide Web Conference, Hrsg. WWW 2012, April 16–20, Lyon: International World Wide Web Conference Committee

Official Google Blog (2012) Introducing Google Play: All your entertainment, anywhere you go, Verfügbar unter: <https://googleblog.blogspot.de/2012/03/introducing-google-play-all-your.html> [abgerufen am: 06.12.17].

Ott, M. et al., (2011) Finding Deceptive Opinion Spam by Any Stretch of the Imagination. In: Association for Computational Linguistics, Hrsg. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland: Association for Computational Linguistics S. 309.

Ott, Myle, Cardie, Claire, Hancock, Jeff (2012) Estimating the Prevalence of Deception in Online Review Communities. In: Committee, International World Wide Web Conference, Hrsg. WWW 2012, April 16–20, Lyon, Frankreich: International World Wide Web Conference Committee

ReviewMeta (2017) How it Works - ReviewMeta Blog, Verfügbar unter: <https://reviewmeta.com/blog/how-it-works/> [abgerufen am: 20.12.2017].

Statista (2016a) Kumulative Anzahl der Empfehlungen auf Yelp weltweit in den Jahren 2009 bis 2016 (in Millionen), Verfügbar unter: <https://de.statista.com/statistik/daten/studie/218866/umfrage/anzahl-der-geposteten-empfehlungen-auf-yelp/> [abgerufen am: 16.01.2018].

Statista (2016b) Umfrage zum Einfluss von Online-Bewertungen in Deutschland 2016,, Verfügbar unter: <https://de.statista.com/statistik/daten/studie/375999/umfrage/einfluss-von-online-bewertungen-nach-geschlecht/> [abgerufen am: 13.12.2017].

test (2010) Jeder ein Tester, 5/2012 S. S. 76-79, Verfügbar unter: <https://www.test.de/Hotelbewertung-Die-besten-Portale-im-Netz-1841156-2841156/> [abgerufen am: 02.01.2017].

test (2012) Meerblick per Mausclick, 5/2012 S. S. 77-80, Verfügbar unter: <https://www.test.de/Hotelbuchung-Meerblick-per-Mausclick-4367789-0/> [abgerufen am: 02.01.2017].

Tiku, Nitasha (2017) FACEBOOK'S NEW CAPTCHA TEST: 'UPLOAD A CLEAR PHOTO OF YOUR FACE', Verfügbar unter: <https://www.wired.com/story/facebooks-new-captcha-test-upload-a-clear-photo-of-your-face/> [abgerufen am: 30.12.2017].

TU Dresden Juniorprofessur für Methoden der Psychologie & Kognitive Modellierung (2017) Quasi-Experimente, Verfügbar unter: <http://versuch.file2.wcms.tu-dresden.de/w/index.php/Quasi-Experimente> [abgerufen am: 12.12.2017].

Wu, G. et al. (2010) Distortion as a validation criterion in the identification of suspicious reviews, Dublin: University College Dublin, Verfügbar unter: <http://www.csi.ucd.ie/content/distortion-validation-criterionidentification-suspicious-reviews> [abgerufen am: 20.12.2017].

Anhang

In der elektronischen Abgabe dieser Arbeit bzw. auf der beigelegten CD ist als Anhang eine Excel-Datei zu finden, in welcher die Planung des Experiments inkl. aller abgegebenen Bewertungen sowie die Messergebnisse zu finden sind.