FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Political Disparity? in Twitter

S e m i n a r B i g D a t a

at the
Friedrisch Schiller University of Jena
Faculty of Mathematics and Computer Science
Graduate Degree Computer Science

submitted to
Prof. Dr. Clemens Beckstein
Schloßgasse 10
07743 Jena

submitted by
Kenny Gozali
Chris Gerlach
and Walter Ehrenberger

Jena, January 27, 2023

Abstract

In der vorliegenden Arbeit behandeln wir eine Sentimentalitätsanalyse von US amerikanischen Politikern aus dem *'House of Representatives'*. Dazu haben wir Daten von Twitter der letzten 12 Jahren zu den genannten Repräsentatnten *gescrapt* und mit Hilfe von des Big Data Frameworks Spark verarbeitet. Ziel der Sentimentalitätsanalyse war es Unterschiede der beiden Parteien (Republikaner und Demokraten) zu bestimmten politischen und auch allgemeinen Themen herauszufiltern. Jedoch haben sich in den gegebenen Daten weniger Diskrepanzen zwischen den beiden Parteien erkennen lassen, als zu Beginn erwartet, wie im Laufe dieser Arbeit verdeutlicht wird.

Contents

List of Figures

1 Introduction

so Text *schreiben* asdfigdfs;oijgs;oef.

sadfasfdas ldfas dlfj asldfnasdfasl

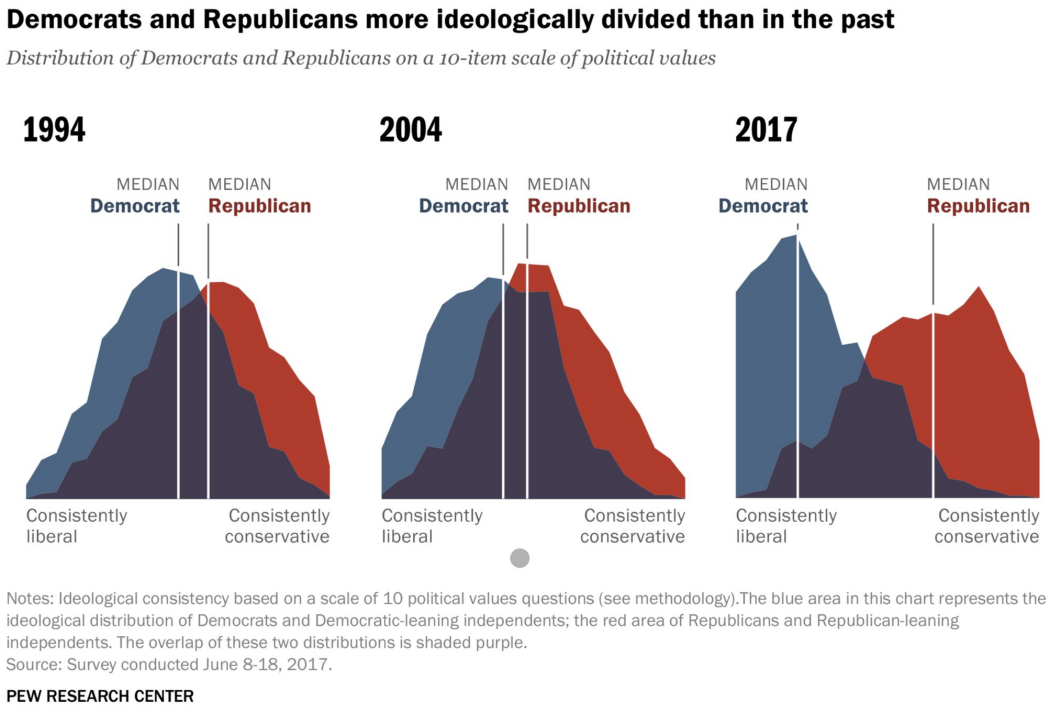

Figure 1.1: Ine Beschreibung f

figureBild

Hier kommt der Text hin.

Zitiere [Moo06] Internetverzeichnis.

Oder zitiere [al.14] Literaturverzeichnis.

2 Getting data (good title missing)

2.1 Background

- Allgemeine Vorstellung der Packages ihr Vor und Nachteile

2.1.1 GetOldTweets3-Pakage

- Was ist der Vorteil der Scraping Bibliothek und Nachteile

GetOldTweets3 ist ein kostenlose Python 3 Packages mit welchen Twitterdaten ohne API-Schlüssel abgerufen werden können. Mit GetOldTweets3 können Sie Tweets mit einer Vielzahl von Suchparametern wie Start-/Enddatum, Benutzername(n), Textabfrage und Referenzortbereich durchsuchen. Außerdem können Sie angeben, welche Tweet-Attribute Sie einbeziehen möchten. Einige Attribute sind: Nutzername, Tweettext, Datum, Retweets und Hashtags.[1] Die offizielle API von Twitter hat eine lästige Zeiteinschränkung, weshalb man keine Tweets älter als eine Woche abrufen kann. Es gibt einige Tools die Zugang zu älteren Tweets anbieten, diese sind jedoch meistens kostenpflichtig. Das Forscher- team hat nach einem andere Tool gesucht die diese Aufgaben übernehmen, wodurch die Wahl auf das Package GetOldTweets3 gefallen ist.[2] Die Analyse des Codes von GetOldTweets3 und die Funktionsweise des Searchthrough Browsers von Twitter zeigt wie das Packages auch an alte Tweets kommt. Grundsätzlich, wenn sie auf Twitter seiten eingeben oder User suchen, startet ein Scroll-Loader. D.h. wenn sie dann weiter nach unten scrollen, bekommen sie immer mehr Tweets zu den Suchbegriffen. Diese ganzen tweets bekommen sie durch Abfragen an einen JSON-Provider.

Quellen bis jetzt: [1]https://andrea-yoss.medium.com/getoldtweets3-830ebb8b2dab , [2]https://pypi.or
[3]https://github.com/Jefferson-Henrique/GetOldTweetspython/blob/master/got3/manager/TweetM

- GetOldTweets zieht sich die die Abgespeicherten Json-Datein zu einem bestimmten nutzer oder anderen Vorgaben, welche dem Package übergeben werden kann - Baut einen User-Agenden, welcher dann die Daten zu der zusammengebauten URL

2.1.2 NLTK-Natural language Toolkit

- Vllt Klären warum wir nicht NLTK verwendet haben, sondern Textblob - Vorteile von Textblob gegenüber NLTK - Nachteile von Nltk

2.1.3 TextBlob

- Textblob: Vllt. herausfinden wie die Berechnung stattgefunden hat - Vorteile von Textblob gegenüber NLTK - Warum haben wir das Tool genutzt.

2.1.4 Sparks

- Aus der Vorlesung Vorteile von Spark finden und einbauen

2.2 Scraping and Sanitization

2.2.1 Scraping

2.2.2 Sanitization

2.3 MapReduce (good title missing)

Hallo Palmoooooo und Walta

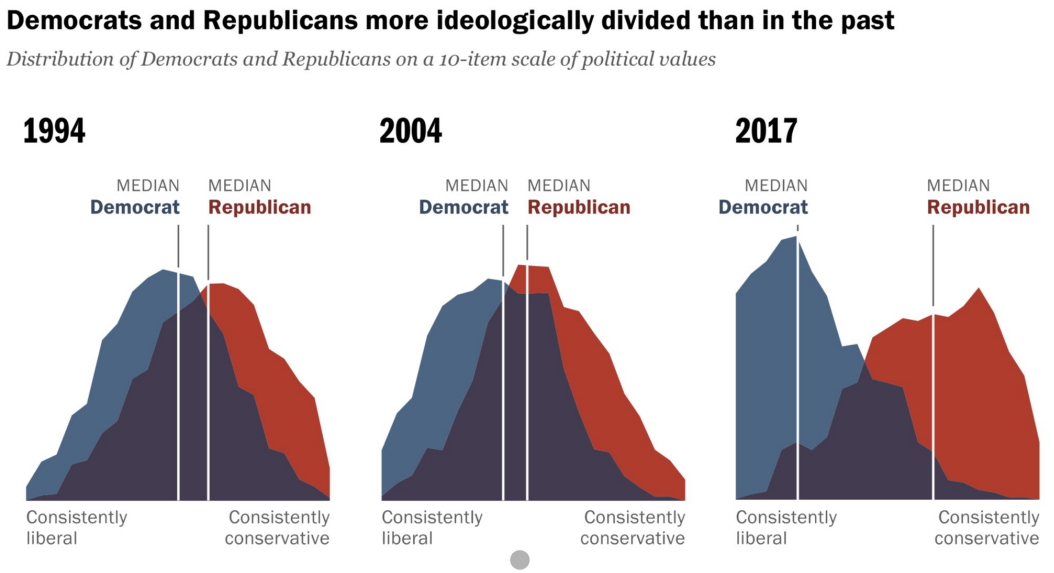

Figure 2.1: Ine Beschreibung f

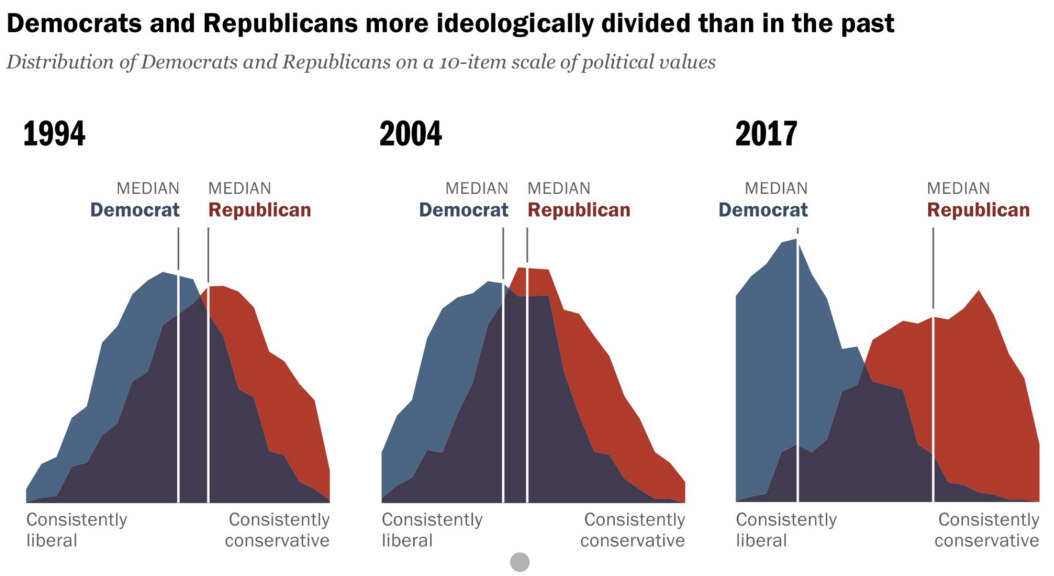

Figure 2.2: Ine Beschreibung f

3 Analysing data (good title missing)

3.1 data1 (good title missing)

Hey Kennyyyyy

3.2 data2 (good title missing)

Hey Kennyyyyy

4 Schluss

4.1 Resume

Hey Kennyyyyy

Bibliography

[al.14] AL., Michael D.: Political Polarization in the American Public. (2014)

Internet sources

[Moo06] Moor, J.: *The Dartmouth College Artificial Intelligence Conference: The Next Fifty years.* 2006. – AI Magazine, Vol 27, No., 4, Pp. 87-9