Wahlforschung mit digitalen Daten

Anwendungsbeispiele und Perspektiven



Glossar

Buzz-Daten

Diese Daten repräsentieren das Volumen an Nennungen für die gesuchten Keywords (Firmen name, Marke, Personenname, Trend, Kampagne) auf Social-Media-Plattformen.

Brandwatch

Eine Social-Listening-Plattform vom gleichnamigen Hersteller, mit der Buzz- und Sentiment-Daten bezogen und analysiert werden können. Brandwatch ist ein langjähriger Partner von TD Reply.

Machine Learning:

Dieser Begriff umfasst IT-Systeme und -Anwendungen, die mithilfe selbstlernender Algorith men aus sehr großen Datensätzen schnell Muster, Gesetzmäßigkeiten und relevante Information extrahieren können.

Social Listening:

Bezieht sich auf Methoden und Tools wie Brandwatch, die zur Extraktion und Analyse von relevanten Informationen aus nutzergenerierten Kommentaren und Diskussionen in sozialen Medien dienen.

Suchdaten:

Unter diesem Begriff werden größtenteils von Suchmaschinen-Anbietern wie Google bereitgestellte Daten verstanden, die Informationen zu gesuchten Begriffen sowie das ihnen zugrundeliegende Suchvolumen in bestimmten Regionen und Zeiträumen beinhalten.

Sentiment-Daten:

Mittels Social-Listening-Tools und Text-Mining-Methoden extrahierte Daten aus Texten in sozialen Medien, die Aufschluss darüber geben, wie positiv oder negativ sich die generelle Haltung (Sentiment) bestimmter Gruppen zu bestimmten Themen darstellt.

Regressions- und Faktorenanalysen

Komplexe Verfahren aus der multivariaten Statistik. Werden oft kombiniert, um viele Variablen in wenige wichtige Faktoren zusammenzuschließen und um Korrelationen zwischen verschie denen Variablen und Faktoren zu identifizieren.

INHALTSVERZEICHNIS

Abstract	01
Einleitung: Neue Wege in der Wahlforschung erkunden	02
Google-Suchen für die Kanzlerkandidaten: Korreliert das Suchinteresse mit Wahlpräferenzen?	05
Mediale Repräsentanz politischer Themen: Was Daten über die Rolle der Medien verraten	08
Prognose der Wahlergebnisse: Medien-Buzz und -Sentiment als Korrektiv für Umfragefehler	13
Wahlbeteiligung: Suchdaten als zuverlässiger Prädiktor	18
Diskussion: Wem nitzt es?	21

Abstract

TD Reply rief Anfang 2021 das Forschungsprojekt "VOICE" ins Leben, das aus früheren Initiativen von TD Reply-Mitarbeitern hervorging. VOICE hatte zum Ziel, auf Basis der Datenexpertise von TD Reply neue, auf digitalen Daten basierende Methoden zur Wahlforschung zu entwickeln - abseits der klassischen umfragebasierten Methoden. Die Validität und Praktikabilität der neuentwickelten Methoden wurde während der Bundestagswahlen 2021 umfangreich getestet. Die wichtigsten gesammelten Erkenntnisse werden in diesem Whitepaper näher beleuchtet.

Zusammengefasst sind die folgenden fünf Haupterkenntnisse des VOICE-Projektes hervorzuheben:

1/

Digitale Daten haben ein hohes erklärendes Potential für Wahlverhalten, das in der heutigen Wahlforschung weitgehend ungenutzt bleibt 3/

Durch Social Listening wird deutlich, dass die Medienrepräsentanz politisch relevanter Themen potentiell einen hohen Einfluss auf das Wahlverhalten ausüben kann

2/

Social Listening, ein aus dem Marketing stammender digitaler Ansatz zur Konsumentenforschung, lässt sich sinnvoll in der Wahlforschung einsetzen 4/

Ein Vorhersagemodell, das Umfragedaten mit Social-Listening-Daten vereint, kann die politische Stimmungslage mit einem Vorlauf von mehreren Wochen prognostizieren

5/

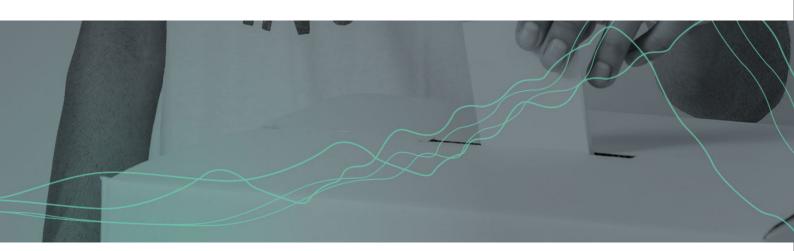
Die Wahlbeteiligung lässt sich mithilfe eines auf Suchdaten basierenden Modells mit einer hohen Genauigkeit vorhersagen

Einleitung:

Neue Wege in der Wahlforschung erkunden

Spätestens seit der Brexit-Wahl in Großbritannien und dem Sieg Donald Trumps bei den US-Präsidentschaftswahlen 2016 wird die Rolle der klassischen Wahlforschung und ihrem Hauptinstrument, der politischen Umfrage, sowohl von gesellschaftlicher als auch wissenschaftlicher Seite zunehmend kritisiert. Auch in Deutschland haben viele Umfragen in den letzten Jahren die Wahlergebnisse insbesondere von vergleichsweise jungen und populistischen Parteien wie der AfD ungenau prognostiziert. So wurde bei den Landtagswahlen 2016 in Sachsen-Anhalt die Stärke der AfD in diesem Bundesland deutlich unterschätzt. Fünf Jahre später, bei den Landtagswahlen 2021, wurde sie wiederum ebenso deutlich überschätzt. Statt dem vielfach prognostizierten Kopf-an-Kopf-Rennen zwischen CDU und AfD gewann die CDU mit einem üppigen Vorsprung von 16 Prozentpunkten.

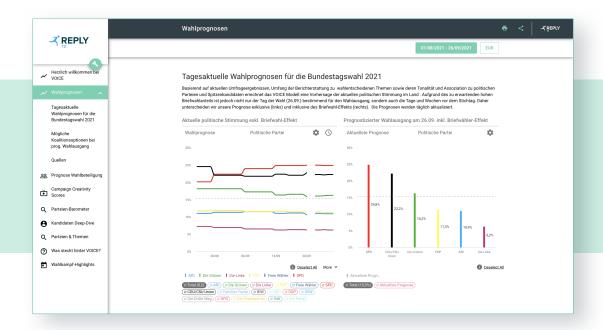
Die Problematik von Umfragen betrifft naturgemäß nicht nur die Wahlforschung, sondern auch die quantitative Marktforschung. Leider setzt auch die recht verkrustete Marktforschung zumindest in Deutschland immer noch auf klassische umfragebasierte Methoden. Allerdings haben Marktforscher, wenn sie denn wollen, längst Zugang zu einer echten und in der Praxis erprobten Alternative



zu Umfragen: der Analyse der Online-Fußspuren von Internetnutzern in Form von Suchbegriffen oder Reaktionen auf Social Media.

TD Reply ist ein führender deutscher Spezialist für diese Art der datengetriebenen Markt- und Konsumverhaltensforschung, die auf dem Prinzip "Zuhören statt Befragen" beruht. Durch die Auswertung von Online-Daten mithilfe datenanalytischer und statistischer Methoden können wir Markttrends mit einer hohen Genauigkeit prognostizieren und ein besseres Verständnis von ihnen bekommen, und das ohne jemandem befragen zu müssen. Das hängt vor allem mit der konkurrenzlos hohen Verfügbarkeit, Granularität und Aktualität von Online-Daten zusammen. Ihr exploratives und vorausschauendes Potential ist enorm. Bereits seit zehn Jahren unterstützen wir erfolgreich mehrere DAX30-Unternehmen mit unseren Datenanalysen.

Vor einiger Zeit ging eine bunt gemischte Gruppe politisch interessierter TD-Reply-Mitarbeiter der Frage nach, ob die gleichen Methoden und Prinzipien auch in der Wahlforschung Anwendung finden könnten. Ihre Untersuchungen sprachen allesamt eine klare Sprache: Mit unseren Ansätzen können wir durchaus nicht nur den Markt und das Konsumverhalten besser verstehen, sondern auch das Wahlverhalten. Damit wurden die Weichen für das Forschungsprojekt "VOICE" gestellt, das im Januar 2021 seinen Anfang nahm. Gegenstand der Untersuchung: das Wahlverhalten der Deutschen bei der Bundestagswahl 2021. Können wir mithilfe digitaler Daten wirklich neue Erkenntnisse über die politischen Präferenzen deutschen Wählerinnen und Wähler gewinnen?



Unser Hauptziel war es, zu zeigen, dass datenbasierten Methoden bei der Analyse und Prognose von Wahlverhalten eine sinnvolle Alternative oder auch Ergänzung zu Umfragen sein können. Nach dem Abschluss des Projektes sehen wir uns darin bestätigt. Nun wollen wir eine Expertendiskussion zu diesem Thema anregen. Dazu dient dieses Whitepaper, das unsere wichtigsten Ergebnisse und Erkenntisse zusammenfasst.

Ebenfalls wichtig für uns ist es deutlich zu machen, dass Big-Data-Analytics-Projekte wie VOICE nicht mit der Nutzung sensibler Personendaten einhergehen müssen. Wir griffen ausschließlich auf anonymisierte und offene Daten zurück.

Letztendlich wollen wir mit den hier gezeigten Einblicken in unser VOICE-Projekt möglichst viele Menschen gerade im Bereich der Wahlforschung dazu inspirieren, neue Wege in der Wahlforschung zu erkunden. Digitale Daten sind dabei die Wegweiser.



Google-Suchen für die Kanzlerkandidaten:

Korreliert das Suchinteresse mit Wahlpräferenzen?

Google-Suchen für die Kanzlerkandidaten:

Korreliert das Suchinteresse mit Wahlpräferenzen?









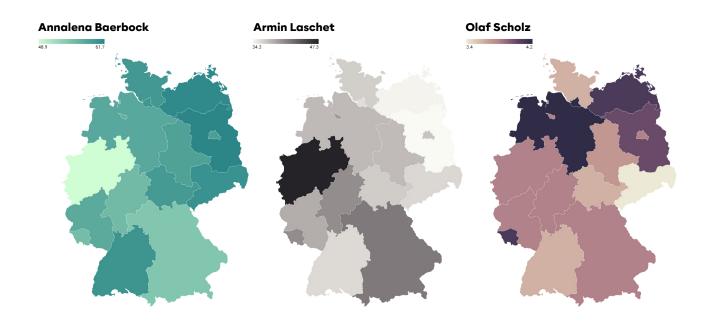
Einleitung

Die Arbeit mit Google-Suchdaten ist ein fester Bestandteil unserer täglichen Leistungen als Marketing- und Innovationsberatung für namhafte kommerzielle Marken. Eine wichtige Metrik in diesem Zusammenhang ist der Anteil am Gesamt-Suchvolumen einer Marke im Vergleich zu anderen Marken in einer bestimmten Kategorie. Das Suchinteresse der Verbraucher ist in vielen Branchen ein zuverlässiger Indikator für künftige Käufe, beispielsweise bei Automodellen. Wird nach einem Automodell im Vergleich zu konkurrierenden Modellen besonders oft gesucht, können wir erfahrungsgemäß eine zuverlässige Prognose zum künftigem Erfolg dieses Modells aufstellen.

Im Rahmen des VOICE-Projektes untersuchten wir mithilfe dieser Metrik, welche Kandidaten während des Wahlkampfes zur Bundestagswahl 2021 das meiste Suchinteresse (Anteil am Suchvolumen) generierten.

Methode

Dafür ersetzten wir Marke durch Kanzlerkandidat und Verbraucher durch Wähler. Berechnet wurde also der Anteil am Gesamt-Suchvolumen pro Kandidat verglichen mit allen relevanten Kandidaten im Bundestagswahlkampf: Annalena Baerbock, Olaf Scholz und Armin Laschet. Die vorteilhafte Eigenschaft von Google-Suchdaten, auf eine regionale Ebene heruntergebrochen werden zu können, erlaubte es uns, das Suchinteresse pro Bundesland zu kalkulieren.



Anteil am Suchvolumen pro Bundesland in Prozent und pro Kanzlerkandidat.

Google-Suchdaten Januar-Juli 2021.

Ergebnis

Im Schnitt erzeugte im Erhebungszeitraum Januar-Juli 2021 Annalena Baerbock das meiste Suchinteresse (56% der Anteile am Suchvolumen), gefolgt von Armin Laschet (40%) und Olaf Scholz (4%). Das ist verglichen mit den endgültigen Wahlergebnissen für die SPD, CDU und die Grünen die genau umgekehrte Reihenfolge. Auffällig bei der regionalen Verteilung ist auch das tendenziell höhere Suchinteresse für Baerbock in den neuen Bundesländern, wo die Grünen das mit Abstand schwächste Standbein haben und auch ihre schlechtesten Ergebnisse erzielten.

Mediale Repräsentanz politischer Themen:

Was Daten über die Rolle der Medien verraten



Mediale Repräsentanz politischer Themen:

Was Daten über die Rolle der Medien verraten

Einleitung

In der Wahlforschung gilt es als gesicherte Erkenntnis, dass die Medienberichterstattung einen gewissen Einfluss auf das Wahlverhalten ausübt. Das ist nicht zuletzt die Lebensgrundlage für Partei-Medienstrategen und PR-Profis. Klar ist, dass Medien in westlichen Demokratien das letzte Wort darüber haben, über welche Themen sie Bericht erstatten (Agenda Setting). Sicher ist auch, dass Medien die Gesprächsthemen unter Familien und Freunden beeinflussen können. Insgesamt herrscht jedoch wenig Konsens in der Wahlforschung über die Art und Stärke der Medien-Effekte. Was können digitale Daten zu der Debate beitragen?

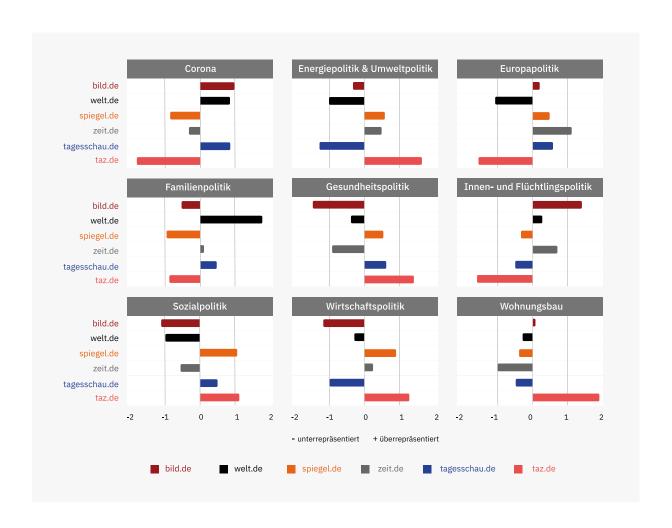
Für das VOICE-Projekt fanden wir nach vielem Experimentieren mit verschiedenen Datenquellen heraus, dass gerade Buzz-Daten mindestens eine robuste quantitative Diskussionbasis schaffen können, die die Forschung voranbringen könnte. Denn sie decken Assoziationen etwa zwischen Thema und Medium (und anderen Variablen) auf, die in ihrer Belastbarkeit, Skalierbarkeit und Verfügbarkeit konkurrenzlos sind. In unserem Fall haben wir auf Grundlage dieser Assoziationen die Repräsentanz verschiedener politischer Themen in führenden Medien errechnet.

Diese sagt auf Basis echter, aktueller Daten aus, wie unter- oder überrepräsentiert ein politisches Thema in einem Medium ist. Dadurch wird auch deutlich, inwiefern gängige Klischees der Wahrheit entsprechen.

Eine vergleichbare Analyse würde ohne die Hilfe digitaler Daten ungleich mehr Zeit und menschliche Ressourcen erfordern oder wäre bei extrem vielen Variablen gar kaum realisierbar.

Methode

Im ersten Schritt wurden innerhalb der Brandwatch-Plattform neun politische Themenfelder samt den jeweils zugehörigen Keywords für jedes Themenfeld definiert. Dann wurde als einzige Datenquelle eine Auswahl der zwanzig reichweitenstärksten deutschsprachigen Online-Zeitungen festgesetzt. Im nächsten Schritt wurde das Buzz-Volumen für diese Themenfelder und diese Medien mittels Brandwatch heruntergeladen. Zeitraum der Datenerhebung war Januar-August 2021. Anschließend wurde auf Basis des Buzz-Volumen-Datensatzes der Anteil am Buzz-Volumen pro Thema und pro Medium kalkuliert. Die relativen Unterschiede zwischen den Medien wurde mittels der statistischen Programmiersprache R visualisiert.



Repräsentanz politischer Themen in führenden Medien (über welche Themen besonders häufig/selten berichtet wird) Buzz-Daten aus der Brandwatch-Plattform, Januar-August 2021.

Ergebnis

Es wurden teilweise enorme Unterschiede in der Themen-Repräsentanz zwischen den verschiedenen Medien festgestellt, die auch recht deutlich die politische Ausrichtung der jeweiligen Medien widerspiegeln. So berichtet beispielsweise die linke taz ungleich häufiger über die Themen "Wohnungsbau", "Energie- und Umweltpolitik" und "Gesundheitspolitik" als die anderen in die Auswahl bezogenen Digital-Medien. Die konservative Welt hingegen besetzt die Themen "Familienpolitik", "Corona" und "Innen- und Flüchtlingspolitik" überproportional, während die taz über diese Themen vergleichsweise wenig Bericht erstattet. Deutlich sichtbar wird auch die starke Überrepräsentanz des Themas "Innen- und Flüchtlingspolitik" bei der Bild, eine der meistbesuchten deutschen Websites. Bei der Tagesschau sind die Themen "Energie- und Umweltpolitik" und "Wirtschaftspolitik" deutlich unterrepräsentiert, während bei anderen Themen die Ausschläge erwartungsgemäß eher gering sind.



MITHILFE VON SOCIAL-LISTENING-DATEN LÄSST SICH DIE MEDIALE THEMEN-REPRÄSENTANZ GENAU QUANTIFIZIEREN.

Diskussion

Die ergebnisseitigen Überraschungen bei dieser Analyse sind überschaubar. Es überrepräsentieren meistens genau die Medien bestimmte Themen, bei denen es zu erwarten war. Und doch zeigt sich hier deutlich, wie der Medieneffekt-Zweig der Wahlforschung von digitalen Daten profitieren kann. Denn "Common Sense" ist keine Wissenschaft und oft genug irreführend. Die Wissenschaft muss sich auf belastbare Zahlen und Fakten stützen, um überzeugende Theorien aufstellen zu können. Digitale Daten können diese erforderliche quantitative Basis mit einem vergleichsweise geringen personellen und zeitlichen Aufwand. Zudem liefert die hier gezeigte Analyse nur einen kurzen Blick auf die Möglichkeiten für die Medieneffekt-Forschung, die mit diesem Social-Listening-Ansatz möglich sind.

Am spannendsten könnte die Kombination mit weiteren Variablen wie Umfragedaten oder Wahlergebnissen in einer Zeitreihe sein. Denn so kann näher untersucht werden, wie die erhöhte Präsenz einiger politischer Themenbereiche in reichweitenstarken Medien die Zustimmungswerte für verschiedene Parteien beeinflusst. Zudem können Sentiment-Daten hinzugezogen werden, um die Tonalität, in der Medien über bestimmte Themen berichten zu untersuchen. Möglich wäre beispielsweise auch die Hinzunahme weiterer Datenquellen oder ein Vergleich der Themen-Repräsentanz zwischen Zeitungen und beliebten Social-Media-Kanälen, um die Rolle von Social Media in der politischen Debatte besser zu verstehen. Den Möglichkeiten für die Medieneffekt-Forschung sind hier kaum Grenzen gesetzt.





Medien-Buzz und -Sentiment als Korrektiv für Umfragefehler

Prognose der Wahlergebnisse:

Medien-Buzz und -Sentiment als Korrektiv für Umfragefehler

Einleitung

Datengetriebene Methoden müssen nicht zwingend im Gegensatz zu klassischen Umfragen stehen, sondern können diese auch sinnvoll ergänzen. Das beweist unser Wahlergebnis-Vorhersagemodell, das wir im Rahmen des VOICE-Projektes für die Prognose der Ergebnisse bei den Bundestagswahlen 2021 entwickelt haben. In diesem Modell wurden aggregierte Umfragedaten mit Social-Listening-Daten kombiniert.

Methode

Im ersten Schritt wurden öffentlich verfügbare aggregierte Umfragedaten abgerufen und für das Modell aufbereitet. Dabei wurde unter anderem der unterschiedliche Bias der Institute in der Bewertung der projizierten Parteienstärke zwischen den Umfrageinstituten herausgerechnet. Um kurzfristigen Ausschlägen nicht zu viel Gewicht bei der Vorhersage beizumessen, wurden die Erhebungsdaten mit einem gleitenden Mittelwert geglättet.

Zudem machten wir mit dem Medien-Sentiment und dem Themen-Buzz von zwei Social-Listening-Kennzahlen Gebrauch, die sich im Vergleich mit anderen Daten als ein besonders verlässliches Korrektiv für die Umfragewerte gezeigt haben. Die Medien-Sentiment-Daten veranschaulichen, in welcher Tonalität Medien über politisch relevante Themen berichten. Der Themen-Buzz spiegelt die mediale Präsenz dieser politischen Themen wider. Diese Daten wurden jeweils von der Brandwatch-Plattform bezogen.

Vorangegangene Bundestagswahlkämpfe haben gezeigt, dass die politische Stimmung zum Wahltag hin eine beschleunigte Dynamik aufweist. Da wir unser Modell auch mit Daten aus der Zeit vor Beginn der Wahlkämpfe trainiert haben um eine ausreichende Datenmenge zu erhalten, wurde dieser Anstieg explizit mitmodelliert.

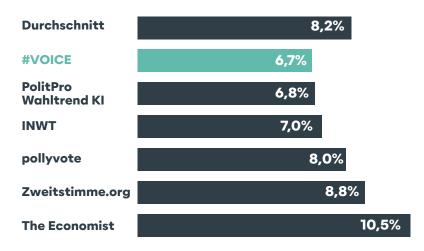
Zudem wurde der Briefwahl-Effekt im Modell explizit berücksichtigt, da aufgrund der Corona-Krise deutlich mehr Briefwähler zu erwarten waren als in den Jahren zuvor. Wir gingen in unserem Modell von einem Briefwähler-Anteil von 50% aus. Außerdem wurde angenommen, dass die Briefwähler zwischen 15. August und dem Wahltag abstimmen, wobei darin jeder Tag als gleichwahrscheinlich für die Stimmabgabe angenommen wurde.

Als Vorhersagemodell wurde ein vektor-autoregressives Modell (VAR-Modell) gewählt.

Ergebnis

Die absolute prozentuale Abweichung vom offiziellen Wahlendergebnis unseres Vorhersagemodells betrug 6,70%. Damit war unsere Prognose auf dem siebten Platz der zehn genauesten Prognosen und noch deutlich genauer als die Prognosen vieler bekannter Umfrageinstitute wie Forsa (7,60%), Kantar (8,60%) und YouGov (10,20%).

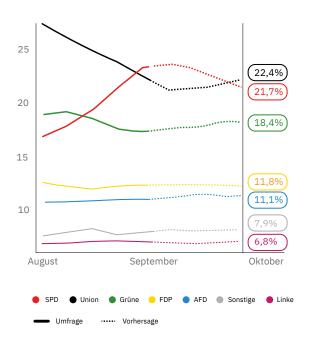
Rangliste: Wer den Wahlausgang am besten prognostizierte

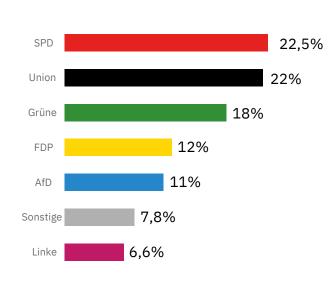


Absolute Gesamtabweichung vom vorläufigen offiziellen Endergebnis der BTW21 nach Aggregator Recherche von TD Reply.

Buzz-Daten aus der Brandwatch-Plattform, Januar-August 2021.

Noch wichtiger ist allerdings, dass unser Modell sich mit der richtigen Prognose allgemeiner Wahltrends von den Umfragen absetzen konnte. Bereits einen Monat vor dem Stichtag hat es eine starke Aufholjagd der Union im Wahlkampf-Endspurt prognostiziert, die sich in der letzten Wahlkampf-Woche tatsächlich manifestierte. Unter Einbeziehung des Briefwahleffekts wurde gleichzeitig prognostiziert, dass die Union es aufgrund des Vorsprunges der SPD dennoch nicht schaffen würde, die stärkste Partei zu werden. Diese Prognose war ebenfalls richtig.





Entwicklung der politischen Stimmung in den Wochen vor der Bundestagswahl 2021 mit prognostizierten Wahlergenissen, Briefwahleffekt nicht einbezogen. Prognostizierter Wahlausgang bei den Bundestagswahlen 2021 unter Einbeziehung des Briefwahleffekts.

Die hier gezeigte Prognose wurde am 02.09.2021, also rund drei Wochen vor dem Wahltag abgegeben und auf den Social-Media-Kanälen von TD Reply publiziert.



Diskussion

Die hohe Genauigkeit unseres Vorhersagemodells ist angesichts der Tatsache, dass unser Vorhersage-Modell von Data Scientists und Marketing-Beratern erstellt wurde, die bislang noch nichts Vergleichbares gemacht haben, beeindruckend. Das beweist den Nutzen von Medien-Buzz- und Sentiment-Daten als Korrektiv für Umfragen. Die in der vorangegangen Fallstudie aufgestellte Vermutung, dass die mediale Berichterstattung einen signifikanten Einfluss auf das Wahlverhalten ausübt, wird damit bestätigt. Diese Effekte sollten in der Wahlforschung mithilfe Medien-Buzz und -Sentiment-Daten genauer untersucht werden.



MITHILFE VON SOCIAL-LISTENING-DATEN ALS
KORREKTIV KONNTEN WIR DIE ENTWICKLUNG DER
POLITISCHEN STIMMUNG PRÄZISE VORHERSAGEN.

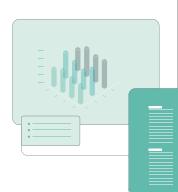
Wahlbeteiligung:

Suchdaten als zuverlässiger Prädiktor



Wahlbeteiligung:

Suchdaten als zuverlässiger Prädiktor



Einleitung

Die Wahlbeteiligung ist ein Gesundheitsindikator für westliche Demokratien. In Deutschland sinkt die Wahlbeteiligung seit vielen Jahrzehnten. Umso wichtiger ist es zu verstehen, welche Faktoren die Wahlbeteiligung beeinflussen.

Unsere Vermutung ist, dass gesteigertes Suchinteresse nach politischen Themen in einer geographischen Region positiv mit einer höheren Wahlbeteiligung in dieser Region einhergeht. Für das VOICE-Projekt haben wir uns darum gewagt, ein auf Suchdaten basierendes Vorhersagemodell für die Wahlbeteiligung bei der Bundestagswahl 2021 zu entwickeln, um diese Hypothese zu testen.

Methode

Ziel war die Untersuchung der Rolle von Google-Suchdaten als erklärende Variable für Veränderungen in der Wahlbeteiligung. Es wurden zunächst zwei Keyword-Cluster erstellt für wahlspezifische (z.B. "wählen gehen") und allgemeine politikrelevante Keywords (z.B. "Bundeskanzler"). Für diese Keywords wurden anschließend via Google Adwords die organischen Google-Suchen ab 2017 (Jahr der vorletzten Bundestagswahl) auf Bundesland-Ebene bezogen.

Neben Google-Suchdaten wurden weitere erklärende Variablen hinzugezogen, um die relative Faktorstärke von Suchdaten einordnen zu können. Dazu gehören vorherige Wahlbeteiligungen, die Entwicklung des Bruttoinlandproduktes (BIP) pro Kopf sowie der sogenannte Herfindahl-Index, mit dem determiniert werden kann, wie gleichmäßig Parteisuchen verteilt sind.

Mittels faktorenanalytischer und clusteranalytischer Verfahren wie PCA und kmeans wurde der Zusammenhang der verschiedenen Variablen mit der Wahlbeteiligung in den untersuchten Regionen überprüft. Die Variablen werden dann in ein auf Random-Forest-Regression basierendes Machine-Learning-Modell über führt, um die Bedeutung der Suchdaten als Prädiktor für die Wahlbeteiligung zu testen. Dieses Modell trifft aus verschiedenen zufälligen Datenkombinationen die mittlere beste Prognose.

Ergebnis

Weil Google die organischen Suchdaten für einen Monat erst ab der Mitte des nächsten Monats verfügbar macht, mussten wir aufgrund des ungünstigen Stichtags unsere erste Vorhersage der Wahlbeteiligung bei der Bundestagswahl 2021 mit den Suchdaten vom August 2021 treffen. Unsere prognostizierte Wahlbeteiligung lag bei 72%, was einer Abweichung von -4,6% (76,6%) vom offiziellen Endergebnis entspricht.

Nachdem wir das Modell aber erneut mit den Suchdaten vom September 2021 laufen ließen, erzielten wir eine prognostizierte Wahlbeiteilung von 76,5% was beinahe auf den Punkt genau der offiziellen Wahlbeteiligung entspricht. Damit wurde unsere Vermutung, dass Suchdaten ein wichtiger Prädiktor für die Wahlbeteiligung sein können, bestätigt. Diese prognostizierte gesamtdeutsche Wahlbeteiligung setzt sich aus allen prognostizierten Wahlbeteiligungen pro Bundesland zusammen.

Reelle Wahlbeteiligung



1. Wahlbeteiligung bei den letzten Wahlen

im Vorhersagemodell

Die wichtigsten Prädiktoren

- 2. Wachstum BIP pro Kopf im Vergleich zum Vorjahr
- 3. Summe aller Suchen im Keyword-Cluster Wählen gehen im Juli-August 2017
- 4. Summe aller Suchen im Keyword-Cluster Wahlen allgemein im Juli-August 2017
- 5. Herfindahl-Index im Juli-August 2017, zeigt an wie gleichmäßig Parteisuchen verteilt sind

Prognositizierte Wahlbeteiligung



Diskussion

Mithilfe aktueller Suchdaten konnte unser Modell die Wahlbeteiligung bei den Bundestagswahlen 2021 beinahe auf den Punkt genau vorhersagen. Das zeigt, dass Google-Suchen heute untrennbar mit dem Wahlverhalten verschmolzen sind. Die Vernachlässigung von Suchdaten in der Wahlforschung ist darum umso bedauernswerter. Das vorausschauende sowie explorative Potential von Suchdaten, gerade bei der Analyse regionaler Unterschiede in der Wahlbeteiligung, bleibt weitgehend unangetastet. Möglicherweise noch interessanter als die Vorhersage selbst ist allerdings ein genauerer Blick auf die zugrundeliegenden Suchdaten. Wonach wird gesucht, und wann wird besonders oft danach gesucht? Woran könnte es liegen? Welche regionalen Unterschiede herrschen bei den Suchpräferenzen?

Durch unsere beschränkten Ressourcen war es uns in diesem Fall nicht möglich, diesen Fragen in der notwendigen Tiefe nachzugehen. Doch hier offenbart sich ein enormes Potential für die Wahlforschung, mithilfe von Suchdaten neue Erkenntnisse zu den Faktoren hinter der Wahlbeteiligung und dem Stand der Demokratie im Land zu gewinnen.



DAS AUF SUCHDATEN BASIERENDE WAHLBETEILIGUNGS-VORHERSAGEMODELL KONNTE DIE WAHLBETEILIGUNG MIT EINER SEHR HOHEN GENAUIGKEIT PROGNOSTIZIEREN.

Diskussion



Diskussion:

Wem nütztes?

Die in diesem Whitepaper vorgestellten und unserem "VOICE"-Projekt entnommenen Anwendungsbeispiele für digitale Daten in der Wahlforschung sollen vor allem als Diskussionsanstoß dienen. Sie machen deutlich, dass Wahlforscher mithilfe von Suchdaten und Buzz-Daten neue Wege in der Erforschung wahlverhaltensrelevanter Phänomene eröffnen können. Im Vergleich zu Umfrage daten sind diese Daten gehaltvoller, leichter verfügbar, günstiger, aktueller, und weniger verfälschbar. Wir wünschen uns eine rege Expertendiskussion in der Wahlforschung zur Nutzung von datengetriebenen Methoden als Alternative und auch als Ergänzung zu umfragebasierten Methoden. Angesichts der einzigartigen Eigenschaften und der Vorteile von digitalen Daten für die Erforschung des Wahlverhaltens sind wir davon überzeugt, dass die entsprechenden Methoden in den nächsten Jahren ihren Platz neben klassischen Umfragen behaupten werden.

Datenbasierte Methoden haben dabei nicht nur einen akademischen Wert als Katalysatoren für neue Erkenntnisse zum Wahlverhalten und dem Stand der Demokratie im Land. Für "First-Mover," also die Organisationen, die datengetriebene Methoden möglichst früh adoptieren, haben sie entscheidende Wettbewerbs-Vorteile zur Folge. Meinungsforschungs-Instituten erlauben sie es, ihren



analytischen Horizont zu erweitern und neue Geschäftsmodelle entwickeln. Parteien und politische Stiftungen können auf Basis datenbasierter Erkenntnisse schneller und fundierter auf Veränderungen in der Stimmungslandschaft und neue Trends reagieren. Nicht zuletzt bringt die Nutzung datengetriebener Methoden auch den Aufbau starker Datenkompetenzen innerhalb der Organisation mit sich. In unserer heutigen "Data Economy" bilden diese Kompetenzen immer mehr das Fundament für nachhaltigen Erfolg.

Mehr noch: Gerade die Corona-Krise hat eindrucksvoll gezeigt, dass die Nutzung digitaler Daten und ein korrekter Umgang mit ihnen immer mehr Teil einer gesellschaftlichen Erwartung und Kompetenz wird. Der von Experten oft kritisierte mangelhafte Umgang mit Daten seitens der Medien und der Politik, der den vielzitierten "Corona-Blindflug" zur Folge hatte, offenbarte zwar einerseits den eklatanten Mangel an Datenkompetenzen in der Gesellschaft. Andererseits wurde eine wichtige Diskussion angestoßen, und die Datenkompetenz der Meinungsmacher, der Regierenden und auch der allgemeinen Gesellschaft im Allgemeinen dürfte angesichts dieser Erfahrungen auf lange Sicht steigen. Im Zusammenspiel mit der durch die Krise deutlich beschleunigten Digitalisierung hat dies zur Folge, dass in unserer Gesellschaft immer mehr digitale Daten generiert werden, die für Erkenntnisgewinne analysiert werden müssen. Gerade für die krisengeplagten aber für unser Staats- und Gemeinwesen so wichtigen klassischen Medien bietet sich die Chance, mittels eines kompetenten Umgangs mit Daten ihr Überleben und ihre Kredibilität langfristig zu sichern.

Unterm Strich ist die Förderung der Datenkompetenzen und die Adoption datengetriebener Methoden für Wahlforscher, die Medien, und die Politik nicht eine Frage von "Ob", sondern von "Wann." Das "Wann" wird entscheidend über den künftigen Erfolg von Meinungsforschungsinstituten, politischen Stiftungen, Parteien und Medien sein.



Mitwirkende und Kontakt

VOICE-Team: Miruna Antonescu, Philipp Ahrens, Philipp Auerbach, Aakash Dharma, Dr. Ilse Klanner, Florian Lück, Stephanie Riedl, Leticia Puerta Mascaro, Sebastian Pimiento, Irina Shumakova, Alice Sweitzer, Phillip Schalk, Marie Wulfmeyer, Felix Zappe, Leonid Zalischiker

Data Science: Stephanie Riedl, Philip Schalk

Redaktion: Leonid Zalischiker **Design:** Leticia Puerta Mascaro

Anfragen und Anregungen gerne per E-Mail an:



Niklas StogPartner, TD Reply
n.stog@reply.de



Dr. Philipp DauderstädtPartner, TD Reply
n.stog@reply.de

Interesse an individuellen Analysen? Wir helfen Ihnen gerne weiter!



Vielen Dank an Brandwatch für die Unterstützung des VOICE-Projekts!



We drive change through data



