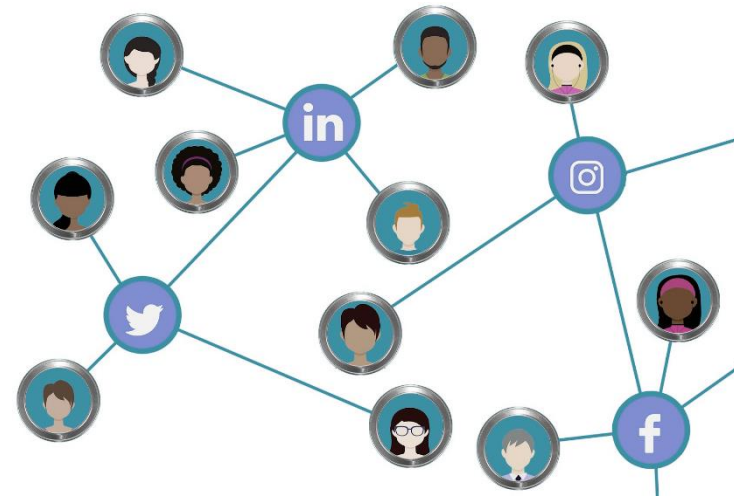


CAS Big Data Analytics

(Big) Data Analytics in Marketing

Hakuna MaData
Lisa-Christina Winter
Co-Founder & Data Scientist
T direkt: +41 79 840 64 34
Mailadresse: lisa@hakuna-madata.com



Vorstellung




Lisa-Christina Winter (@lisachwinter  )

Psychologie (Prom. 2017, Universität Graz)

Statistik & Mathematische Modellierung, Kommunikation, Social Media

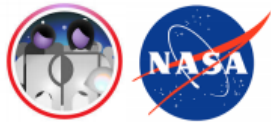
Programmierkenntnisse: R, IBM SPSS

Ehemals: Data Scientist/Consultant @  **ELCA**  **gateB**

Aktuell: Senior User Researcher @  **TestingTime**

Co-Founder @ **Hakuna MaData**

(Growthhacking & Data Science @hakunamadacom  )



Zum Ausprobieren...

Praxisbeispiele

- Webcrawling (legal)
 - <https://hakuna-madata.com/luziNET/liste.php>
- Downloads R & R Studio
 - <https://cran.r-project.org>
 - <https://www.rstudio.com>
- R-Skripte & Daten:
 - <https://www.hakuna-madata.com/hslu.zip>

Marketing: eine Definition

- Der Managementprozess, durch welchen Güter und Services sich vom Konzept zum Kunden bewegen. Marketing inkludiert die Koordination von 4 Elementen (die **4 P's** des Marketing):

(1) Identifikation, Selektion und Entwicklung eines **P**roduktes

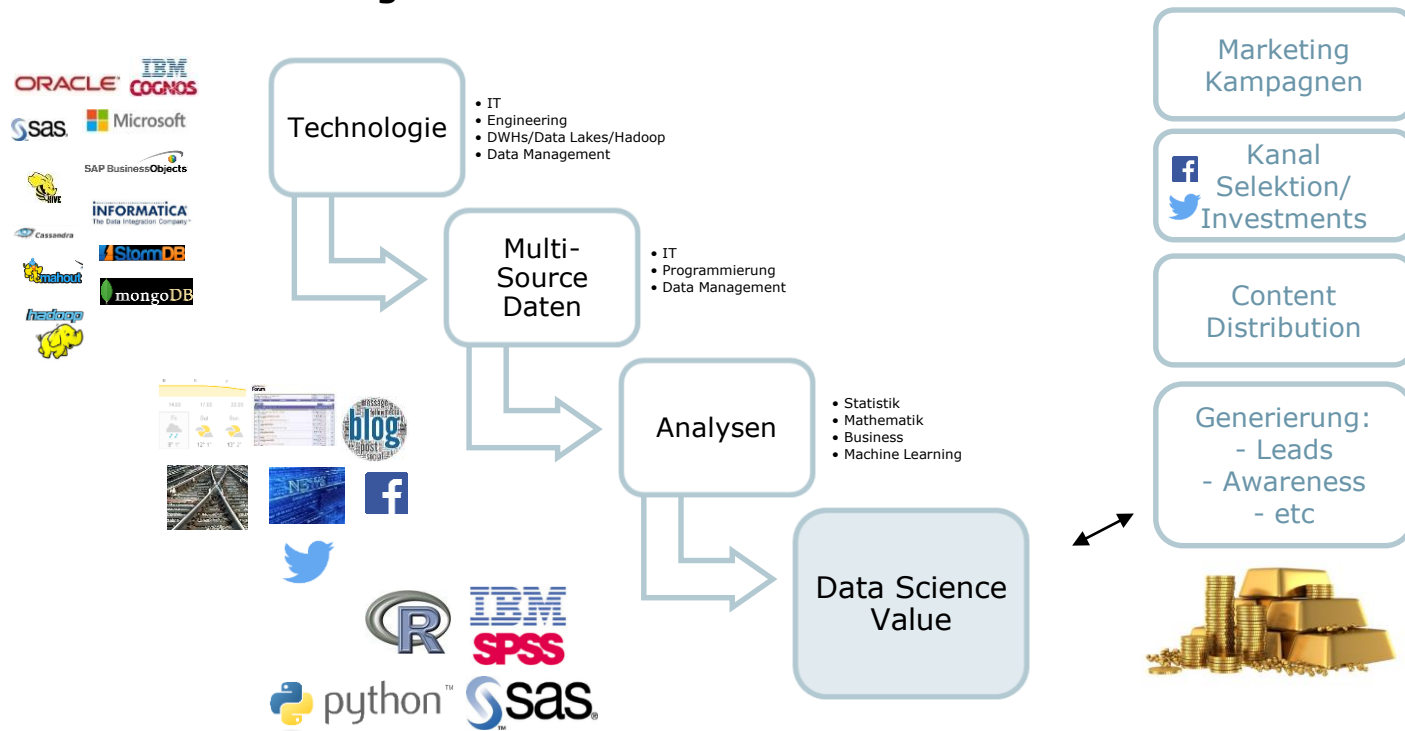
(2) Bestimmung des **P**reises (+ Markt)

(3) Selektion des Distributionskanals → **P**latz des Kunden (+ Channel)

(4) Entwicklung und Implementierung der **P**romotionsstrategie (+ Darstellung & Werbebotschaft)

<http://www.businessdictionary.com/definition/marketing.html>

Big Data – Von der Technologie zum Mehrwert



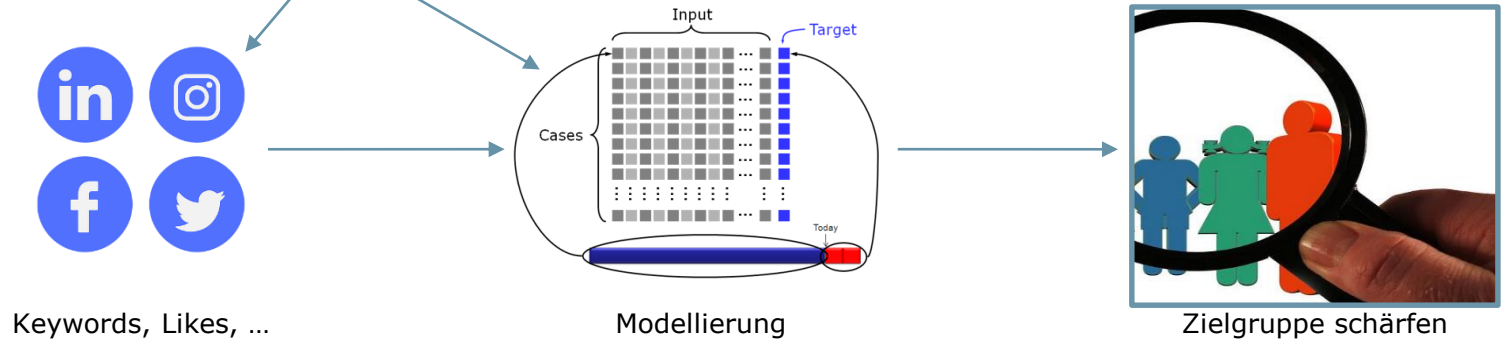
(Big) Data Science im Marketing – Use Cases

Predictive Analytics/Affinitätsanalysen

Ausgangslage



Vorgehen



Von unstrukturierten zu strukturierten Daten ("Data Lake" → DWH)

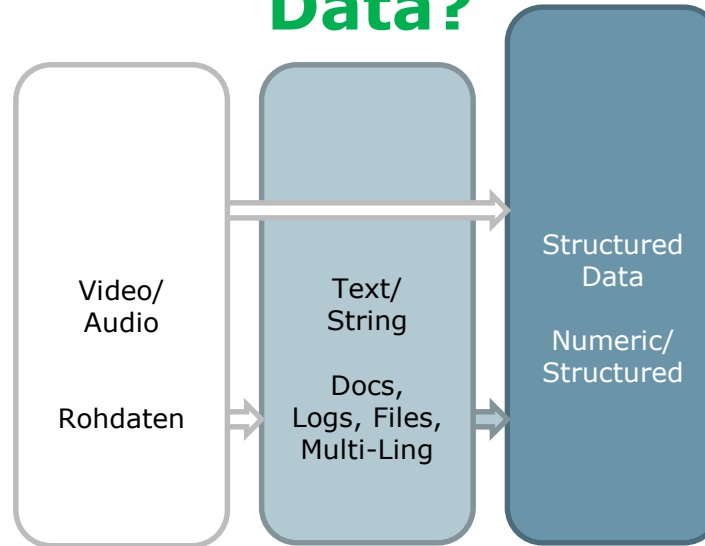
Unstrukturierte → Strukturierte Daten

Text Analytics, Natural Language Processing (NLP) and Search

- Advanced search and monitoring
- Named entity recognition
- Key-phrase/topic extraction
- Sentiment analysis
- Speech recognition
- Question answering
- Automatic translation
- ...

Big Data?

✓volume, ✓velocity, ✓variety, ✓veracity

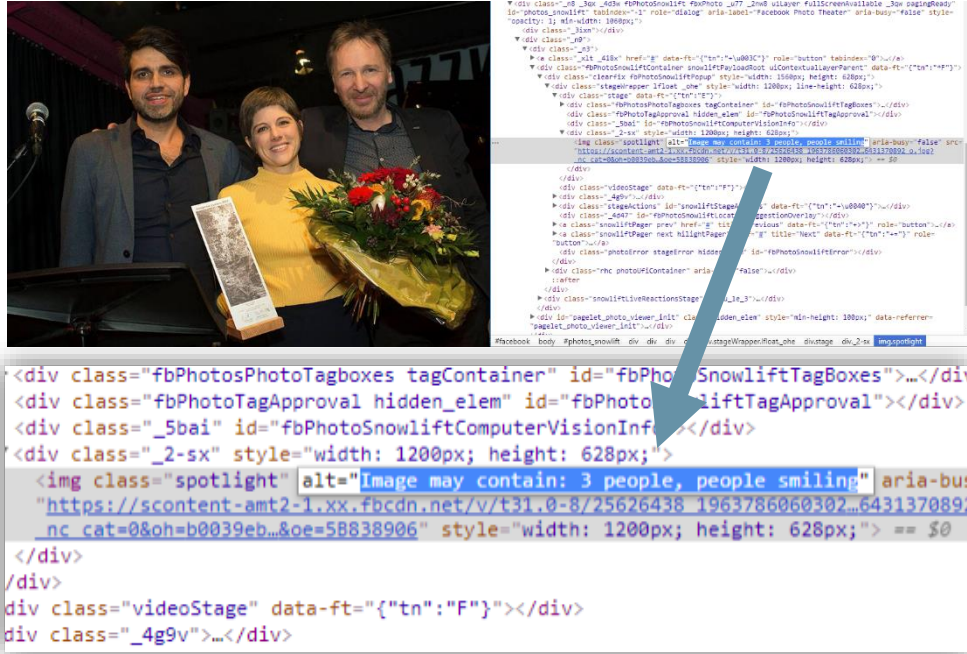


unstrukturiert → strukturiert

Advanced Statistics & Mathematical Modeling

- Predictive Modeling/Data Mining
- Machine Learning/Inference Statistics/Statistical Modeling
- Clustering/Classification
- Visualisation/Exploration
- Model Assessment/Comparison
- Model Scoring
- Bayes/Monte Carlo Simulations
- Discrete mathematical models
- ...

Von unstrukturierten zu strukturierten Daten ("Data Lake" → DWH)



Screenshot: <https://www.facebook.com/hslu.ch/photos/a.266362740045281.87212.227326347282254/1963786060302932/>

Advanced Analytics (Bild & Text)

Text Analytics: Natural-Language Processing

- Natural-Language Processing (**NLP**) ist ein breites Gebiet in der Informatik
- **Spracherkennung** sowie **computergestütztes** Verständnis/Erzeugen „natürlicher Sprache“
- Wörter nicht mehr nur isoliert, sondern im **grammatikalischen Kontext** (Syntax) analysiert
- **Semantischen Aspekte:**
 - Zeichenerkennung (Optical character recognition, OCR)
 - Sentiment Analysis
 - Computergestützte Übersetzung
 - Erzeugung „natürlicher“ Sprache

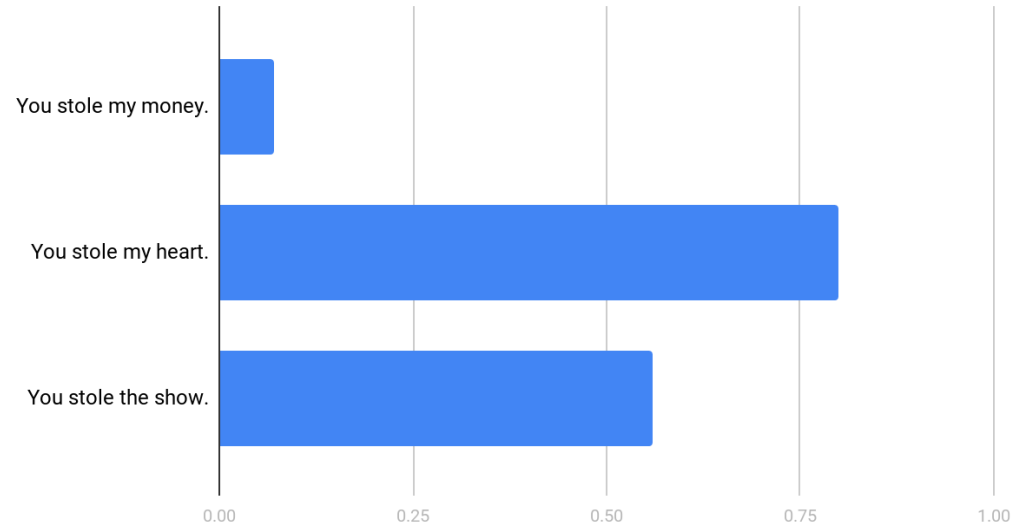
Advanced Social Media Analytics (Bild & Text)

Text Analytics: Sentiment Analysis

- Bietet u.a. die Möglichkeit, Texte hinsichtlich ihrer „**Positivität**“ einzuschätzen
- Rechts sind drei kurze Sätze nach ihrer „Positivität“ eingeschätzt

<https://indico.io/product>
(Service kürzlich eingestellt)

Positivity



Advanced Analytics (Bild & Text)

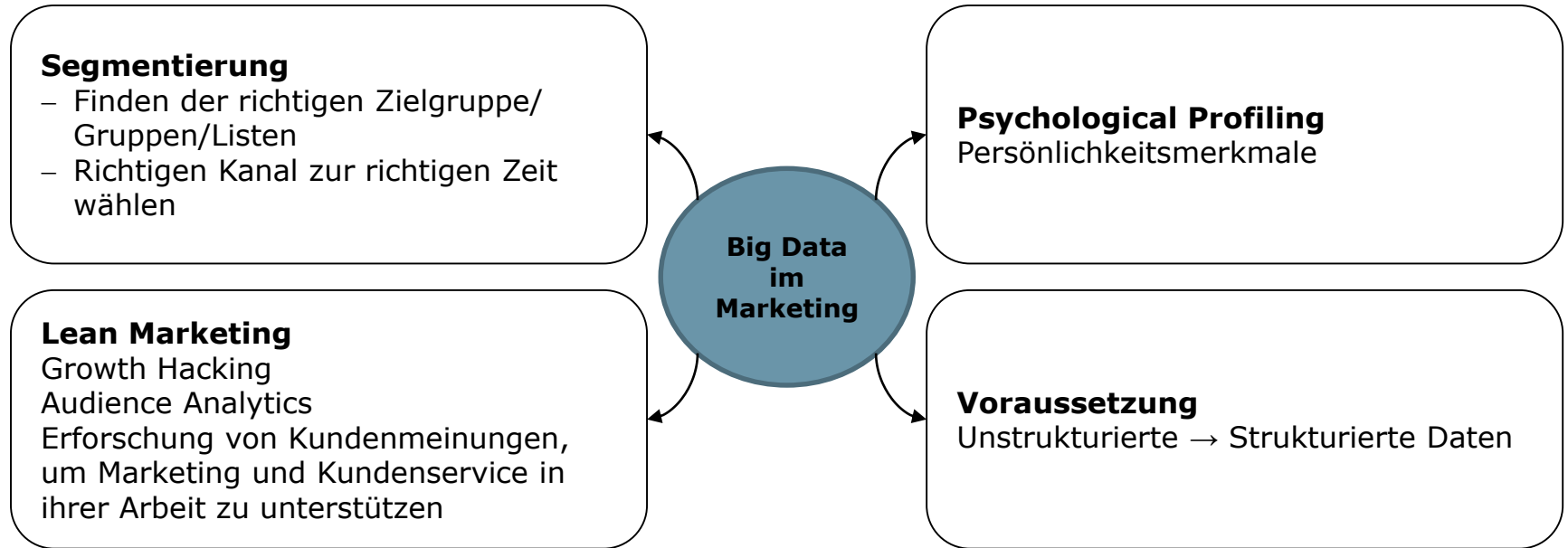
Anwendungsbereiche

- Aus Bilder können viele Informationen extrahiert werden
- Diese Informationen können für **Marketingzwecke** (mit und ohne CRM) verwendet werden
- **REAL TIME!** Post mit einem Bild von einem Baum → Aufruf für eine Spende zum Thema Wald
- Viele Posts mit Bilder von Autos → Werbung mit Autoversicherung
- Informationen im CRM ablegen und Modelle mit entsprechender Affinität auf Kundenbestand erstellen (Direktmarketing)
- Angereicherte Informationen als einfache Trigger im Direktmarketing verwenden (ohne Modelle)

(Big) Data Science im Marketing – Use Cases

(Big) Data Science im Marketing – Use Cases

Social Media Analytics – Anwendung



(Big) Data Science im Marketing – Use Cases

Psychological Profiling

Kombination von bekannten Daten (CRM) & Persönlichkeitsdaten

- z.B. Big Five (Offenheit, Gewissenhaftigkeit, Extraversion, Verträglichkeit, Neurotizismus)
- Identifiziert auf Basis von „behavioral indicators“ auf Social Media

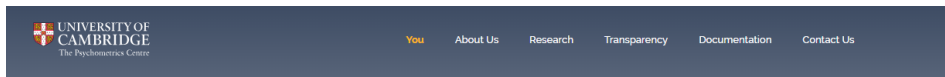


Zum Beispiel: Cambridge University <https://applymagicsauce.com>

The Psychology of Data Science <https://thegroupofanalysts.com/2018/04/03/the-psychology-of-data-science>

(Big) Data Science im Marketing – Use Cases

Psychological Profiling

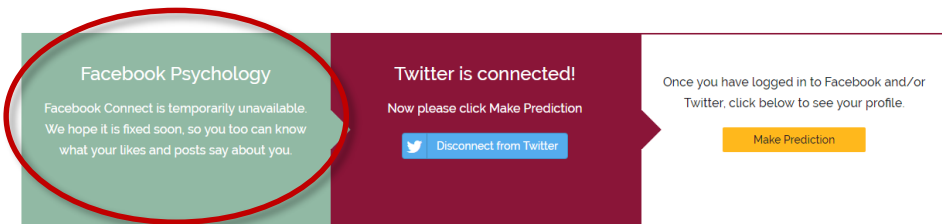


Facebook and Twitter Prediction

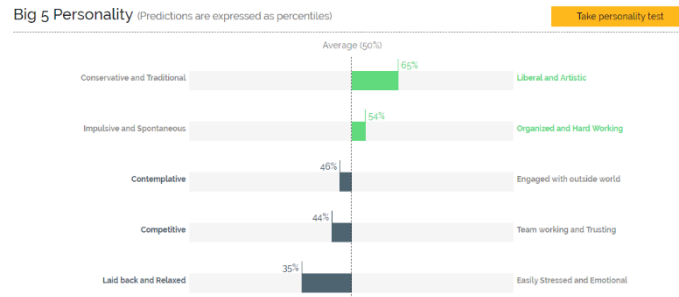
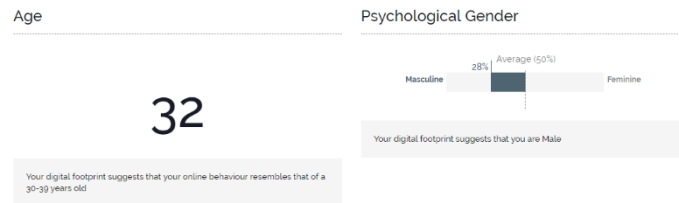
This tool predicts your psycho-demographic profile from digital footprints of your behaviour. It reveals how you might be perceived by others online and provides detailed insights on your personality, intelligence, life satisfaction and more.

Predictions are based on opt-in psychological ground truth from over 6 million volunteers, and our data has been used in over 45 peer-reviewed scientific articles.

Connect your Facebook and Twitter profiles using the log in buttons below to discover what your data says about you. Your use of this demo is anonymous, we will never post on your wall and the results of your prediction will not be stored. We hope you enjoy it!



Cambridge University <https://applymagicsauce.com>



(Big) Data Science im Marketing – Use Cases

Targeted Marketing

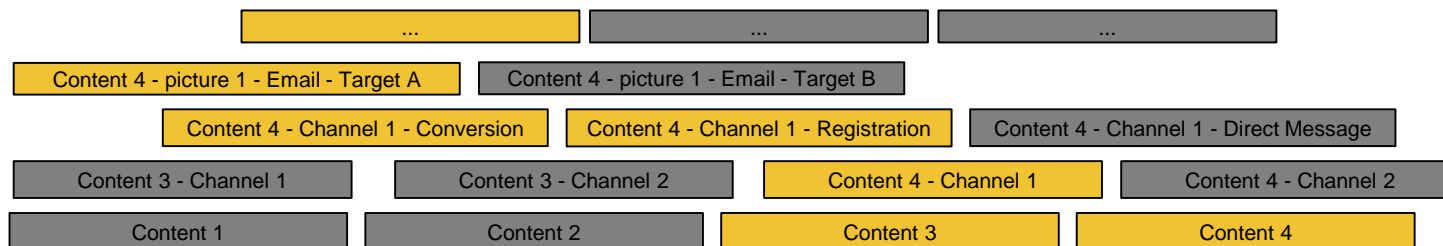
- Typische **Zielvorstellungen** sind
 - Umsatzsteigerung
 - das Einholen von Einschätzungen der eigenen Produkte und Dienstleistungen
 - Verbesserung der öffentlichen Meinung über bestimmte Produkte und/oder Unternehmensaktivitäten
 - Verbessertes **Targeting** (Werbung)
- Identifizierung und Umwandlung der **profitabelsten Kundensegmente**
- Kundensegmente **priorisieren** und Marketinginvestitionen entsprechend anpassen



(Big) Data Science im Marketing – Use Cases

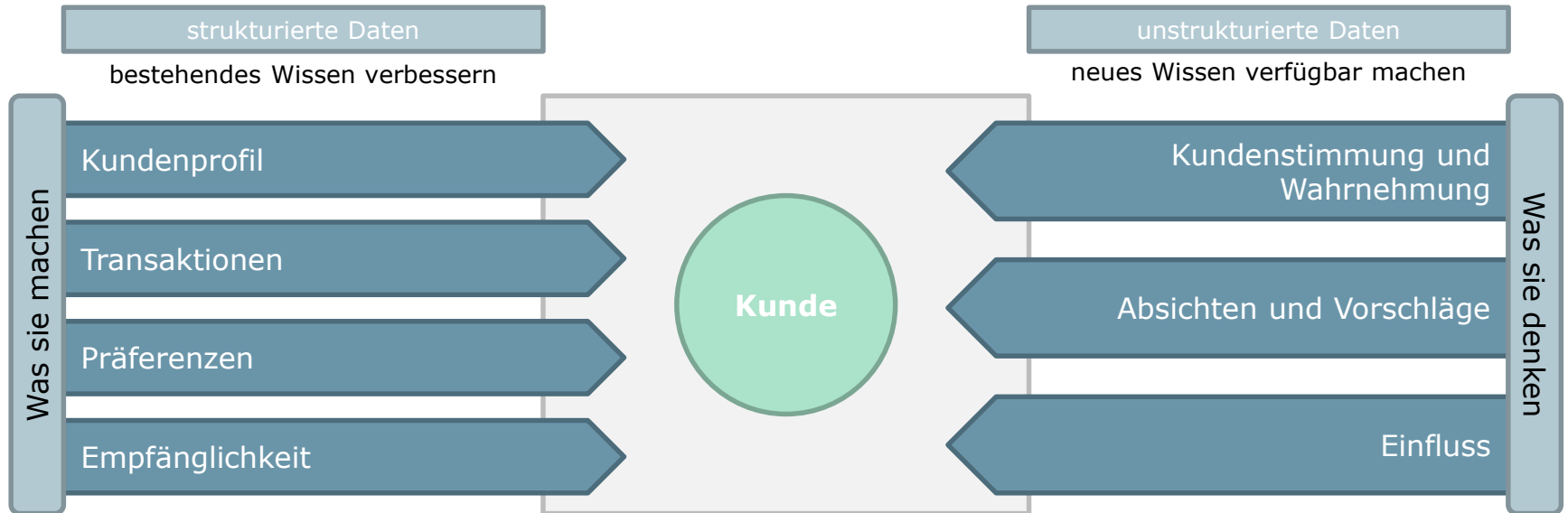
Growth Hacking

- **Growth Hacking** (Experimental Marketing)
- Ziel: Wachstum von **Startups** (Clients & Revenue)
- Geringe finanzielle Mittel (Marketingbudget)
- Lean, creative, experimental Marketing auf spezielle Zielgruppen bezogen
- **Data Driven**: Tools (APIs), Web-Scraping, Social Media Daten aus Portalen



(Big) Data Science im Marketing – Use Cases

CRM & Big Data– Zusammenspiel von unstrukturierten und strukturierten Daten



Affinitätsanalysen/Anreicherung CRM/Predictive Analytics

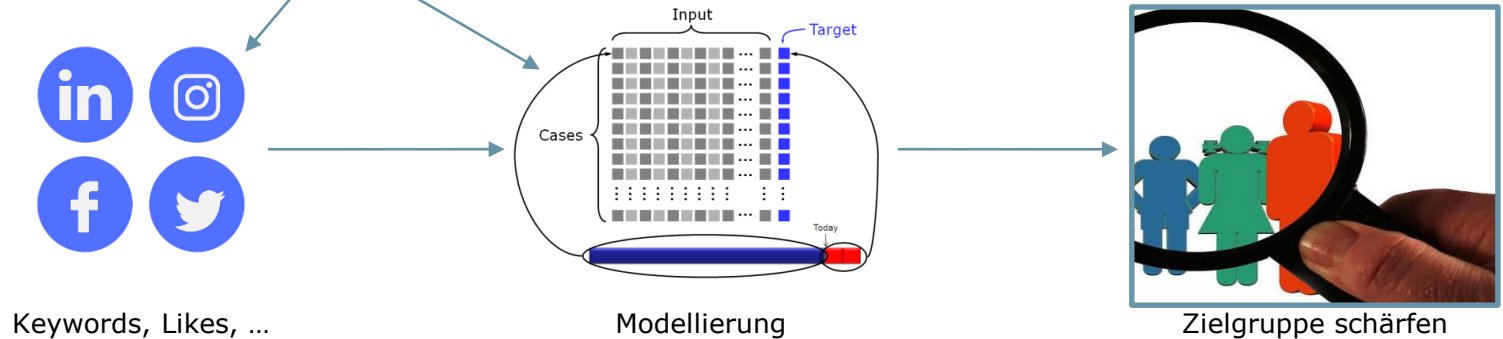
(Big) Data Science im Marketing – Use Cases

Reminder: Predictive Analytics/Affinitätsanalysen

Ausgangslage



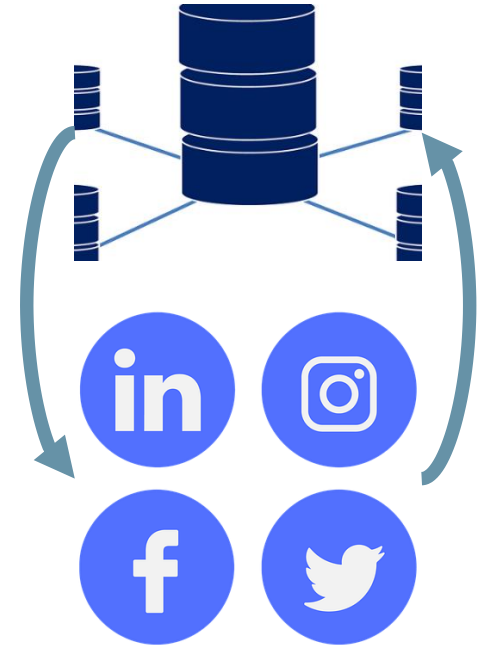
Vorgehen



Affinitätsanalysen/Anreicherung CRM

Ausgangslage

- Vorhandenes **CRM** System (MS Dynamics, SAP-CRM, Eigensysteme, ...)
- Kundenbestand mit:
 - Vor-, Nachname, Adresse
 - E-Mail, Telefon
 - Alter
 - Kaufverhalten (Transaktionen)
 - **Fehlende Informationen zu Soziodemographie**
 - **Fehlende Informationen zu Interessen**



Affinitätsanalysen/Anreicherung CRM

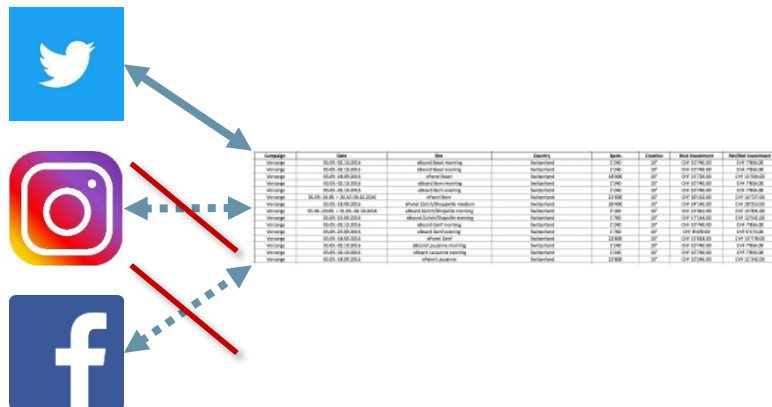
Ziele

- **Kundenbestand:** im CRM mit Daten/Personen aus Social Media Kanäle finden und ausbauen
- **Extrahieren:** von Keywords, Likes, Kategorien, Standort, ... durch Bild- und Texterkennung
- **Anbindung:** ans CRM (Informationen im CRM ablegen)
- **Zielgruppe:** bei Selektionen schärfen (personalisierte Werbeaktionen)
- **Erhöhung:** des Engagements in Bezug auf spezifische Themen durch Affinitätsanalyse

Affinitätsanalysen/Anreicherung CRM

Finden von Personen aufgrund von Name oder Standort/Beruf/...

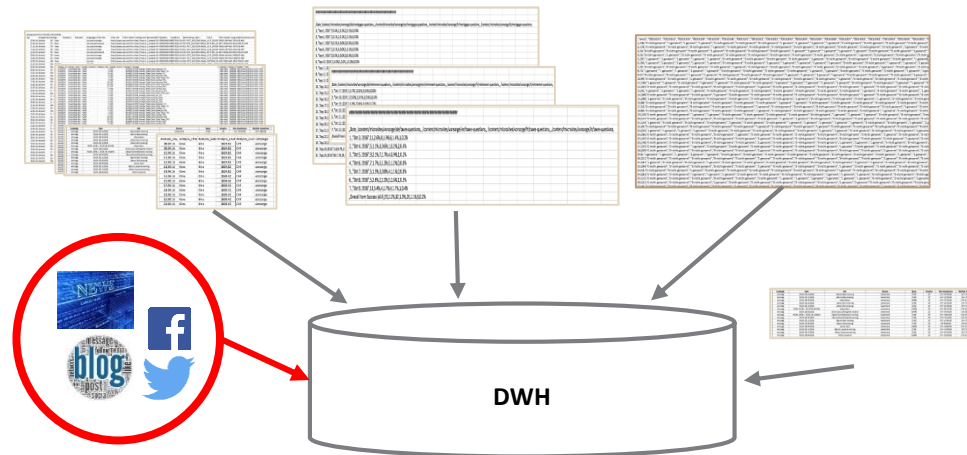
- Auf Social Media können über **APIs** diverse Identifikations-Informationen von Personen abgerufen werden (E-Mail, Adresse, ...), sofern diese freigegeben wurden
- Mit einfachen **Matching-Algorithmen** können diese mit dem Kundenbestand im CRM verknüpft werden
- In der Vergangenheit konnte man Personen eindeutig über **Telefonnummern oder E-Mail-Adressen** finden, was derzeit nicht möglich ist
- Facebook/Instagram haben ihre **APIs massiv eingeschränkt**
- Idealerweise hinterlassen Kunden ihre **Usernamen** in den sozialen Netzwerken im CRM
- Damit erhält man letztlich Informationen über die Personen, indem man sich z.B. ihre Tweets und anderen Interaktionen ansieht



Affinitätsanalysen/Anreicherung CRM

Extrahieren von Keywords/Likes/Kategorien/Standort

- Auf Social Media können über API diverse Identifikations-Informationen von Personen oder Tweets abgerufen werden, sofern diese freigegeben wurde



Affinitätsanalysen/Anreicherung CRM

Praxisbeispiel Affinitätsanalyse



Ausgangslage Firma XY

- Die Firma XY möchte ein **neues Produkt** lancieren: **Einhorn Plüschtier**
- Sie kennen die Zielgruppe, haben aber auf ihrem Kundenbestand keine sinnvollen Informationen, um eine vernünftige Selektion durchzuführen

Ziele

- Um zu schauen, wie das Produkt bei den Kunden ankommt, wollen sie als Test **nur Kunden** angehen, die für das neue Produkt am **affinsten** sind
- Kundenbestand im CRM mit **Daten/Personen aus Twitter finden** und zuordnen
- Extrahieren von **Hashtag #unicorn** aus Twitter
- Selektion der Kunden mit zwei Varianten
 - Nur die Kunden, welche durch Twitter über den Hashtag identifiziert wurden
 - Nachteil:** Alle Kunden, die nicht mit Twitter verknüpft werden konnten, aber affin sind, gehen verloren
 - Modelle zur Identifizierung mit Affinitätsscores (Information von Unicorn vorhanden)
 - Vorteil:** Potential für Selektion wird erhöht

Affinitätsanalysen/Anreicherung CRM

Praxisbeispiel: Affinitätsanalyse

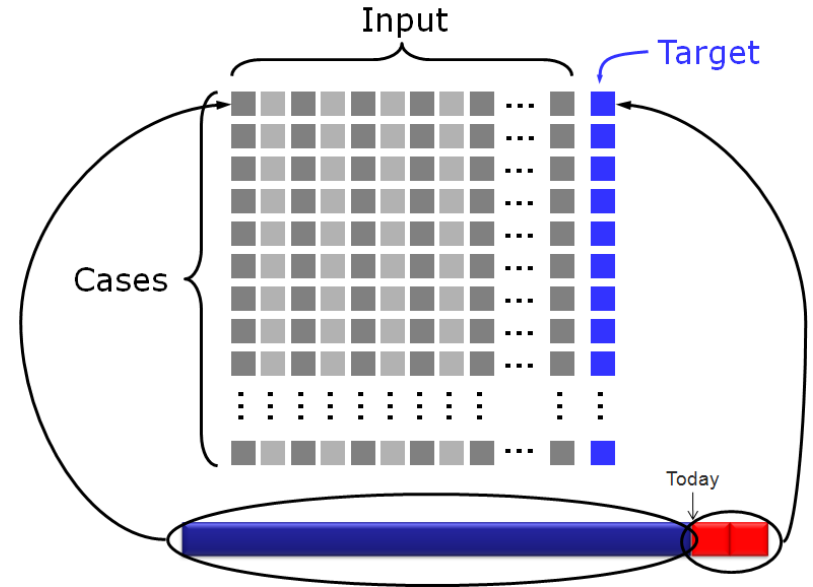
Datenaufbereitung → Analysebasistabelle

Input:

- (Sozio-)Demographie
- Beziehungen und Kontakte (CRM)
- Aktuelle und vergangene Käufe
 - Kaufhäufigkeit und -Beträge
 - Topics of Interest (Keywords)
- Teilnahmen an Aktionen
 - Topics of Interest
- 3rd Party Daten
- Regionale Merkmale

Target:

- Affinität Produkt ja/nein, während des „Performance-Zeitraums“

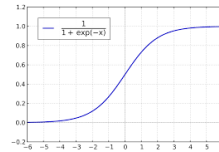


Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

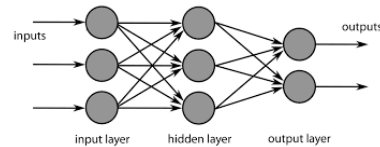
Praxisbeispiel: Affinitätsanalyse

Häufigste Modellierungsansätze für Binäre Zielvariable

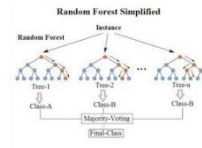
- Logistische Regression



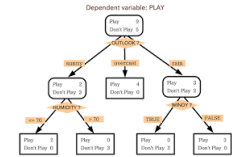
- Neuronale Netze



- Random Forests



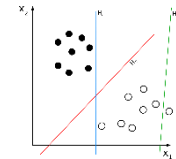
- Entscheidungsbäume



- Naive Bayes



- Support Vector Machines



Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Szenario

(Fiktive!) Daten aus der Sales-Abteilung zum Verkauf eines neuen Produkts: **Einhorn Plüschtier**

Aus Social Media wurde, wie in den Zielen beschrieben, Informationen aus Twitter extrahiert, wer Interesse an Einhörner hat (Zielvariable).

400 Beobachtungen auf 4 Variablen:

- **target (y)**: „0“ (kein Hashtag mit Unicorn gefunden) oder „1“ (Hashtag mit Unicorn gefunden)
 - **gender**: „female“, „male“
 - **hipster**: „no“, „yes“
 - **works_startup**: „no“, „yes“
- } Input

Die unabhängigen Variablen (Input) gender, hipster und works_startup wurden durch Analysen auf sozialen Medien erhoben.

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Modellvorhersagen

Was lernen wir aus dem Modell?

Alle Erkenntnisse beziehen sich auf die Wahrscheinlichkeit für das Produkt, wenn der Hashtag vorhanden ist:

- Männer und Frauen **unterscheiden sich nicht**
- Personen, die sich als „**Hipster**“ beschreiben, haben eine höhere Wahrscheinlichkeit (Odds = 5.9x höher, $p < .001$)
- Den **stärksten Effekt** gibt es bei Personen, die angeben, dass sie in **Startups** arbeiten. Bei diesen sind die Odds, das Produkt zu kaufen 28.9x höher ($p < .001$)

NB: Die Daten sind simuliert!

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Anwendung Modellaffinität

Tabelle mit allen Kunden und folgenden Informationen:

Affinitäts-Wahrscheinlichkeit (in %):
für jeden Kunden persönlich gemäss Modelloutput

Perzentile:
eine Einteilung der Wahrscheinlichkeiten (Scores von 1–100) zur vereinfachten Selektion

Kundeninformationen:
Alle verfügbaren Kundeninformationen als weitere Selektions- und Filtermerkmale

Kunde	Segment	Affinität	Perzentil
Z. Alpha	A	75%	1
Y. Beta	A	65%	1
X. Delta	C	42%	7
W. Omega	B	24%	50
V. Psi	C	14%	69
T. Gamma	B	5%	90
U. Lambda	A	1%	100

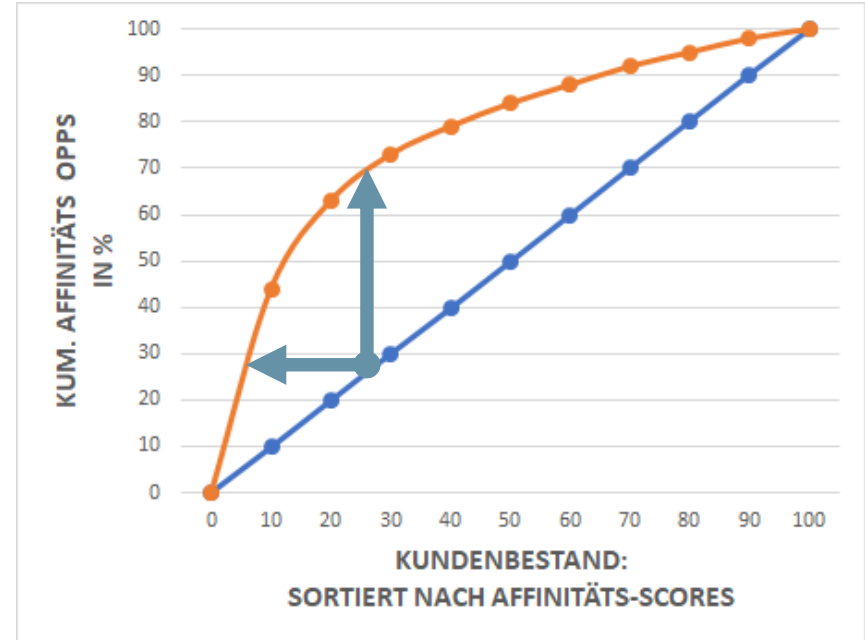
Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Evaluation/Beurteilung Modell

Durch die Modelle können wesentliche Optimierungen erreicht werden:

- Innerhalb der 25% der affinsten Kunden werden dank dem analytischen Modell 70% der Käufer gefunden (**Uplift**)!
- Mit gleichem Aufwand (Kampagnenbudget) lässt sich ein **besseres Ergebnis** bei spezifischer Kundenzielgruppe erreichen
- Alternative Betrachtung des Uplifts: Mit **weniger Aufwand** lässt sich **dasselbe Ergebnis** (dieselbe Kundenloyalisierung) erreichen



Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

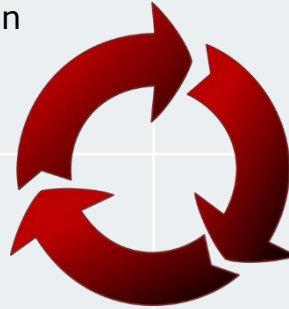
Ausblick

Das Vorgehen kann auf
beliebig viele Keywords/Kategorien
angewendet werden

Es können auch **Cluster-Ansätze**
verwendet werden

Keywords/Kategorien können
miteinander **kombiniert** werden

Infos aus
unterschiedlichen Quellen
verwenden,
um genügend Daten für die
Modellierung zu haben



(Big) Data Science im Marketing – Use Cases

API und Restriktionen („APIcalypse now!“)

- Durch die **Einschränkungen** von Facebook (und Instagram) im Zuge der **Cambridge Analytica** Affäre kaum noch die Möglichkeit, Personen automatisiert eindeutig zu finden
- **Personenbezogene Informationen** kann man jetzt
 - manuell suchen (oder ggf. automatisiert mit händischer Validierung/halb-manuell z.B. Chrome Extensions)
 - durch die freiwillige Angabe der Personen erhalten
 - ...
- **DSGVO (GDPR)** schränkt die Möglichkeiten, sensible personenbezogene Daten zu verarbeiten ein (einige Unternehmen und Anbieter wollen das Risiko eines Verstosses nicht eingehen)
- **Öffentliche Informationen** wie Tweets etc. sind weiterhin zugänglich
- **Web-Scraping** ist oft ein Graubereich bzw. ist fallweise in Nutzungsbedingungen geregelt

(Big) Data Science im Marketing – Use Cases

Ethische Überlegungen

Der Fall von Cambridge Analytica

- **Persönlichkeitstest** (über App) mit 270'000 **Facebook-Usern** durchgeführt und Persönlichkeitsprofil abgebildet
- Berechtigung für die App erlaubte auch das Auslesen persönlicher Daten der **Facebook-Kontakte der Testpersonen**
- Mit etwa 200 Facebook-Freunden pro Testperson sind das rund 50 Millionen Nutzer insgesamt
- **Methode** entwickelt, um mittels „Likes“ auf Facebook ein Persönlichkeitsprofil zu rekonstruieren
- CA erstellte damit die Persönlichkeitsprofile von über 100 Millionen eingetragenen Wählerinnen und Wählern in den USA, um dann unterschiedliche Wahlkampfanzeigen auf FB anzuzeigen
- Die **Datensegmentierung und -analyse** durch CA bedeutet eine radikale Veränderung, wie Technologie verwendet werden kann, um z.B. politisch Einfluss zu nehmen

Schlusswort

- **360-Grad Kundensicht** aufbauen
- **Ausbau vom Kundenbestand** durch Anbindung weiterer Quellen
- Kundenbindung erhöhen durch zusätzliche Informationen aus der Social Media Welt und gezieltes Marketing
- **Neue Modellierungsmöglichkeiten** durch Ergänzung von Online-Daten im CRM
- **Empfehlungen in Real Time** durch aktuelle Online-Informationen
- Reputation Management in Real Time
- Unmittelbares "Feedback" was Veränderungen bei Preis, Produktpalette, Kampagnen etc. anbelangt
- Marketing Mix Optimierung beinahe in Echtzeit:
 - Erhöhung oder Reduktion von Investitionen in gewissen Kanälen
 - Mathematische Optimierung vom Marketing Mix Model
- Targeted Online Navigation
- Next best activity in Real Time



Appendix: Praktische Beispiele

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Deskriptive Analyse

Eine einfache Auszählung ergibt folgendes:

```
> with(daten, ftable(gender, hipster,  
works_startup, y))
```

			y	0	1
gender	hipster	works_startup			
female	no	no	46	4	
		yes	8	42	
	yes	no	27	23	
		yes	1	49	
male	no	no	41	9	
		yes	10	40	
	yes	no	25	25	
		yes	3	47	

Prozentuell (auf die Spalten bezogen):

```
> 100 * prop.table(with(daten, ftable(gender,  
hipster, works_startup, y)), 1)
```

			y	0	1
gender	hipster	works_startup			
female	no	no	92	8	
		yes	16	84	
	yes	no	54	46	
		yes	2	98	
male	no	no	82	18	
		yes	20	80	
	yes	no	50	50	
		yes	6	94	

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Deskriptive Analyse

Informationen der vorherigen Folie zusammengefasst (Häufigkeiten und Prozent für jede Kombination der unabhängigen Variablen).

(Prozentwerte in „ohne Hashtag“ und „mit Hashtag“ sind redundant, weil jeweils 100% minus die andere Variable)

hip?	Startup?	Frauen		Männer	
		ohne Hashtag	mit Hashtag	ohne Hashtag	mit Hashtag
no	no	46 (92%)	4 (8%)	41 (82%)	9 (18%)
	yes	8 (16%)	42 (84%)	10 (20%)	40 (80%)
yes	no	27 (54%)	23 (46%)	25 (50%)	25 (50%)
	yes	1 (2%)	49 (98%)	3 (6%)	47 (94%)

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Modellselektion

Ein **volles Modell mit allen Interaktionen** beschreibt die Daten gut, lässt sich aber vl. vereinfachen.

```
> modF <- glm(y ~ gender * hipster * works_startup, daten, family = binomial)
> summary(modF)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.44235	0.52129	-4.685	2.80e-06 ***
gendermale	0.92600	0.63815	1.451	0.146763
hipsteryes	2.28200	0.59351	3.845	0.000121 ***
works_startupy	4.10058	0.64850	6.323	2.56e-10 ***
gendermale:hipsteryes	-0.76566	0.75350	-1.016	0.309563
gendermale:works_startupy	-1.19793	0.82526	-1.452	0.146616
hipsteryes:works_startupy	-0.04841	1.23343	-0.039	0.968691
gendermale:hipsteryes:works_startupy	-0.10269	1.48878	-0.069	0.945006

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Modellselektion

Volles Modell in Objekt modS. **Modellselektion** mit drop1() mittels Likelihood-Ratio-Test (LRT)

```
> modS <- modF
> drop1(modS, test = "LRT")
Single term deletions
```

Model:

```
y ~ gender * hipster * works_startup
```

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		339.83	355.83		
gender:hipster:works_startup	1	339.84	353.84	0.004785	0.9449

Ohne Dreifachinteraktion ist der Fit nicht signifikant schlechter ($p = .945$), also Term entfernen und drop1() wiederholen.

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Modellselektion

Terme mit `update()` entfernen anstatt immer das ganze Modell angeben.

Neuerliche Testung zeigt, dass keiner der Terme $< .05$ ist.

Entfernen von `hipster:works_startup` mit dem grössten p -Wert.

```
> modS <- update(modS, . ~ . - gender:hipster:works_startup)
```

```
> drop1(modS, test = "LRT")
```

Model:

```
y ~ gender + hipster + works_startup + gender:hipster + gender:works_startup +  
  hipster:works_startup
```

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		339.84	353.84		
gender:hipster	1	341.40	353.40	1.5576	0.2120
gender:works_startup	1	343.27	355.27	3.4305	0.0640 .
hipster:works_startup	1	339.87	351.87	0.0293	0.8641

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Modellselektion

Modellselektion bis kein Term mehr $< .05$ ist.

Modell enthält nur noch hipster und works_startup (nicht signifikant schlechter als das volle Modell mit 8 Parametern; $p = .573$)

```
> summary(modS)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.8711	0.2661	-7.031	2.04e-12	***
hipsteryes	1.7774	0.3049	5.828	5.59e-09	***
works_startupyess	3.3645	0.3225	10.432	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Affinitätsanalysen/Anreicherung CRM/Predictive Social Media

Praxisbeispiel: Affinitätsanalyse

Logistische Regression: Modellvorhersagen

Vorhersagen für neue Daten mit `predict()` (Default: Logit-transformierte Werte wie in `pred_link`; `type = "response"` berechnet Wahrscheinlichkeiten)

```
> pred_mat <- unique(daten[,c("hipster", "works_startup")])
> pred_mat$pred_link <- predict(modS, newdata = pred_mat)
> pred_mat$pred_prob <- predict(modS, newdata = pred_mat, type = "response")
> pred_mat
```

	hipster	works_startup	pred_link	pred_prob
1	no	no	-1.87111362	0.1334129
101	yes	no	-0.09372022	0.4765871
201	no	yes	1.49339355	0.8165871
301	yes	yes	3.27078695	0.9634129

```
> 100 * prop.table(with(daten, ftable(hipster, works_startup, y)), 1)
```

		y	
		0	1
hipster	works_startup		
no	no	87	13
	yes	18	82
yes	no	52	48
	yes	4	96

Die Werte aus der Vorhersage decken sich mit den tatsächlichen sehr gut (siehe Verbindungslinien).