

# Untersuchung von Gentrifizierung am Beispiel Berlin mittels Big Data Analytics

Kolloquium zur Masterarbeit von Dennis Helweg

Hamburg, 21.12.2018

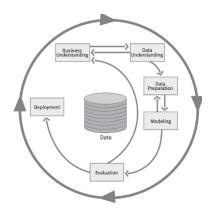
Erstgutachter: Professor Dr. Stefan Voß

Zweitgutachter: Dr. Robert Stahlbock

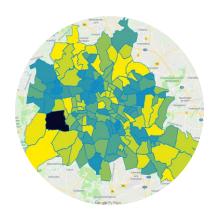




Motivation, Ziel der Arbeit und Forschungsfragen



Theoretische Grundlagen und methodisches Vorgehen



Diskussion der Ergebnisse und Ausblick

# Das Thema Gentrifizierung bewegt die Gemüter – Berlin gilt als Epizentrum der Verdrängung sozial schwacher Mieter durch Aufwertung der Nachbarschaften.



### Motivation und Dimensionen der Gentrifizierung

"In Berlin steigen die Immobilienpreise weltweit am stärksten"

(Berliner Morgenpost, 2018)

"Die kleinen Läden sterben – der Gleimkiez verliert seine Seele" (BZ Berlin, 2018) "8,7 Prozent Rendite – so macht der Immobilienboom auch Sie reich" (Welt, 2018)



"Mietpreisexplosion - Mieter zahlen nach Modernisierung bis zu 200 Prozent mehr" (Focus online, 2017)

"Modernisierung verdrängt Mieter" (WDR, 2018)

"13.000 protestieren gegen steigende Mieten"

(ZEIT ONLINE, 2018)

"Monopoly in Prenzlauer Berg Was geschieht, wenn die **Mittelschicht verdrängt** wird?" (Berliner Zeitung, 2017)

### Die Masterarbeit nutzt allgemein gehaltene Forschungsfragen, die in Abhängigkeit zur Datenlage um Domänenspezifische Fragestellungen erweitert wurden.



### Ziel der Arbeit und Forschungsfragen

Ziel: Untersuchung der Gentrifizierung in Berlin mit Big Data Analytics

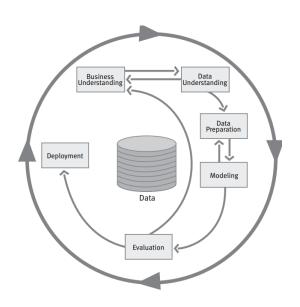
Fokus: Zusammenhang der lokalen Angebotsstruktur mit dem sozialem Status



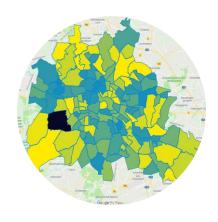




Motivation, Ziel der Arbeit und Forschungsfragen



Theoretische Grundlagen und methodisches Vorgehen



Diskussion der Ergebnisse und Ausblick

### Big Data wird häufig über die 3 V's beschrieben: Volume, Variety und Velocity. Bei der Analyse von Big Data mit Raumbezug sind einige Besonderheiten zu beachten.



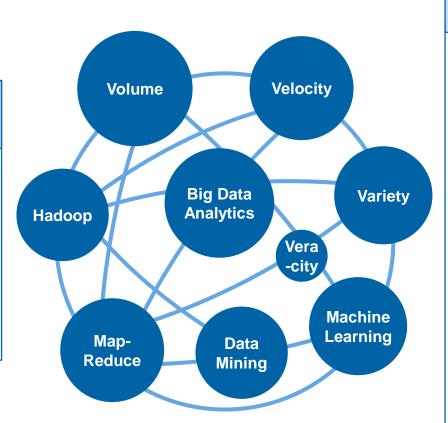
### **Grundlagen Big Data Analytics**

### **Big Data Analytics**Definition



"Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis."

(Gantz & Reinsel, 2011, S. 6)



### **Spatial Big Data**

Big Data mit Raumbezug / Georeferenzierung



### Beispiele:

- Geo-Tags an Bildern
- Standorte an Tweets
- Positionen von POIs
- Standorte(-verläufe) von Mobiltelefonen, ...

### **Spatial Autocorrelation**

"Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

→ Häufung georeferenzierte Bilder bei Instagram

### **Spatial Heterogenity**

- Verschiedene Räumen → Verschiedene Darstellung
  - → Def. Transferleistungsbeziehern an der Bevölkerung

### **Modifiable Area Unit Problem (MAUP)**

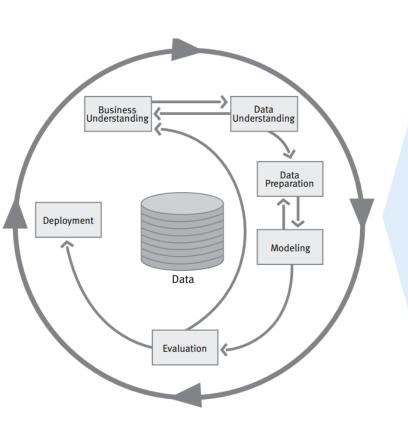
- Daten in verschiedenen Skalenniveaus
- → Raumbezug Zensusdaten vs. Raumbezug KFZ-Zul.
- ullet Verschiedene Auflösungen ullet Verschiedene Ergebnisse
- → Anteil Transferleistungsbeziehern an der Bevölkerung

# Die Methodik der Masterarbeit orientiert sich an dem CRISP-DM Referenzmodell, dem am häufigsten verwendeten Vorgehensmodell bei Big Data Analytics Projekten.



### **Methodik**





- Business Understanding
  Grundlagen Gentrifizierung, Stand der Forschung
- Data Understanding
  Analyse und Bewertung von Datenquellen zur Analyse, Erstellung Hypothesen
- Data PreparationTechnisches System und Datenintegration
- 4 Modeling
  Aufbau von Modellen zu den Forschungsfragen und Hypothesen
- **Evaluation**Evaluation von Datenbasis und Algorithmen, Hypothesen-Überprüfung
- DeploymentDiskussion, Fazit, Ausblick

### Gentrifizierung ist vielschichtig – Es ist ein multidimensionales Phänomen, dass als Prozess in unterschiedliche Phasen aufgeteilt werden kann.

Antell

In %



### **Business Understanding – Grundlagen Gentrifizierung**

#### **Definitionen**

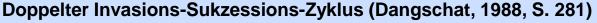
"[Gentrifizierung] beschreibt den Wechsel von einer statusniedrigeren zu einer statushöheren (finanzkräftigeren) **Bewohnerschaft**, der oft mit einer **baulichen Aufwertung**, Veränderungen Eigentümerstruktur und steigenden Mietpreisen einhergeht."

(Deutsches Institut für Urbanistik, 2011)

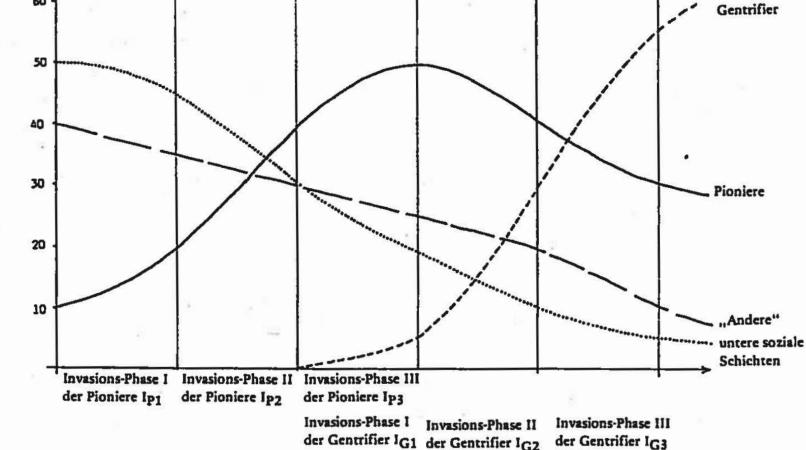
"Gentrification – broadly defined as a **transition** in neighborhood character associated with the displacement of lower-income residents by higher-income newcomers" (Zook, Shelton und Poorthuis, 2017, S. 2)

defining characteristics contemporary gentrification should include in the widest sense: (1) reinvestment of capital; social upgrading of locale by incoming high-income groups; (3) landscape change; and (4) direct or indirect displacement of lowincome groups."

(Davidson und Lees (2005, S. 1170)







## Gentrifizierung wird häufig anhand von Einzelschicksalen betrachtet. Neuere Studien versuchen das Thema mittels Big Data Analytics zu betrachten, jedoch noch nicht in Berlin.



**Business Understanding – Stand der Forschung** 

Big Data Analysen zu Gentrifizierung							
Autor	Ort	Bestimmung Gentrifizierung	Big Data Analytics Ansatz				
Beekmans (2011)	Amsterdam	qualitative Bestimmung	Foursquare POIs und Checkins				
Schaefer (2014)	Los Angeles	qualitative Bestimmung	Twitter Tweets (08.2013-01.2014)				
Venerandi et al. (2015)	London, Birmingham, Manchester	IMD Status (2011)	Foursquare POIs und OSM POIs (04.2014 / 05.2014)				
Hristova et al. (2016)	London	IMD Verbesserung (2010 - 2015)	Geosoziales Netzwerk basierend auf Foursquare Restaurant-POIs und Checkins, sowie Twitter Tweets (12.2010-09.2011)				
Zukin et al. (2015)	Zwei Teile von Brooklyn, New York	qualitative Bestimmung	Yelp Restaurantreviews, manuelle/ qualitative Analyse (05.2014)				

### Gentrifizierung in Berlin



- Bisher keine Big Data Analysen
- Primär qualitative Untersuchungen anhand von Interviews von Betroffenen
- Erste quantitative Studien zu Messung von Gentrifizierung
  - Holm & Schulz (2016) GentriMap
  - Döring & Ulbricht (2016)

Angebotsmietpreise in EUR	Q2_2008 (Top 3)	Ø jähr Anstieg 20 (Differ gesamt	008-2012 enz zu	Q2_2012 (Top 3)	Anstieg (Diffe	Ø jährlicher Anstieg 2012-2016 (Differenz zu gesamt Berlin)	
Bezirk Treptow-Köpenick	5,49	-0,8%		6,48	-0,7%		7,73
Bezirk Marzahn-Hellersdorf	4,73	-1,7%		5,40	-1,2%		6,31
Bezirk Pankow	6,05		0,5%	7,50		0,1%	9,29
Bezirk Neukölln	5,18		1,4%	6,67		1,3%	8,56
Bezirk Charlottenburg-Wilmersdorf	6,78	-0,5%		8,11	-0,4%		9,80
Bezirk Steglitz-Zehlendorf	6,42	-0,9%		7,55	-0,9%		8,96
Bezirk Lichtenberg	5,29		0,3%	6,51		0,1%	7,99
Bezirk Tempelhof-Schöneberg	5,92		0,1%	7,23	-0,1%		8,75
Bezirk Friedrichshain-Kreuzberg	6,11		1,2%	7,80		1,5%	10,14
Bezirk Reinickendorf	5,42	-0,5%		6,47	-0,7%		7,70
Bezirk Spandau	5,22	-1,0%		6,12	-0,8%		7,29
Bezirk Mitte	5,82		1,5%	7,51		1,1%	9,70
Berlin	5,70	5,1	%	6,94	5	,5%	8,52

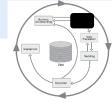
# Für jede Dimension der Gentrifizierung wurden potentielle Datenquellen untersucht. Auf Basis von verschieden Merkmalen wurden drei Datenquellen für die Analyse ausgewählt.



### **Data Understanding – Übersicht Datenquellen**



1



Raumbezug Berlin

Lebensweltlich orientierte

Räume (LOR)

447 Planungsräume

138 Bezirksregionen

60 Prognoseräume

12 Bezirke

1 Berlin

Bezirk 02 Fri.hain-Kreuzberg Prognoseraum 03 Kreuzberg Ost Bezirksregion 04 südl. Luisenstadt

Planungsraum 02 Wrangelkiez



Dimension	Bezeichnung	Kommer- ziell?	Historie?	Daten sammeln / Analytics Lizenz?	Raumbezug	Datenherkunft	Technische Details	Genutzt?
	Mikrozensus	Nein	Ja	Open Data	Bezirke	Öffentlich per Stichprobe	Webinterface mit CSV Export	Nein
Verdrängung und soziale Aufwertung	Außenwanderung	Nein	Ja	Open Data	Bezirke	Meldedaten	Webinterface mit CSV Export	Nein
	Einwohnerregister	Nein	Ja	Open Data	Planungs- räume	Meldedaten	CSV Download	Ja
	Monitoring Soziale Stadtentwicklung	Nein	Ja	Open Data	Planungs- räume	Melde- und Sozialdaten	Excel Download	Ja
	Bestand KFZ	Nein	Ja	Open Data	Gesamt Berlin	Anmeldedaten KFZ	PDF Download	Nein
Immobilien- wirtschaftliche Aufwertung	ImmobilienScout24	Ja	Ja	24h Löschfrist / Keine Angabe	IS24 Räume	Angebote auf Plattform	per API, keine Raumpolygone	Nein
	Google Places API	Ja	Nein	Verbot / Verbot	Koordinaten	Unternehmer / User	per API, 20-60 Treffer pro Aufruf	Nein
Veränderung	Foursquare	Ja	Ja (Create date)	24h Löschfrist / mit Enterprise Liz.	Koordinaten	Unternehmer / User	per API, 50 Treffer pro Aufruf	Nein
der Angebots-	Tripadvisor	Ja	Nein	24h Löschfrist / Verbot	Koordinaten	Unternehmer / User	per API, keine Suche möglich	Nein
struktur	Yelp	Ja	Nein	Erlaubt / Erlaubt	Koordinaten	Unternehmer / User	per API, 1.000 Treffer pro Aufruf	Nein
	OpenStreetMap	Nein	Ja (Create date)	Open Data	Koordinaten	Mapper- Community	Datenbank Download	Ja

### Für die Datenaufbereitung wurden drei Datenquellen und eine eigens entwickelte Mapping-Hierarchie verwendet. Letztere soll zur Aufbereitung der OSM-Daten dienen.



**Data Understanding – Genutzte Datenquellen** 



# 1

#### EINWOHNERREGISTER

- Jährlich verfügbar
- Je Planungsraum Anzahl der Einwohner in Kategorien
  - Alterskategorien (Altersklassen)
  - Geschlecht
  - Migrationshintergrund
  - Ausländer
  - Herkunftsgebiet
  - Wohndauer
  - Wohnlage



### MONITORING SOZIALE STADTENTWICKLUNG

- Alle zwei Jahre verfügbar
- Je Planungsraum Anzahl der Einwohner in Kategorien
- Status- und Dynamik-Index
- Kennzahlen
  - Arbeitslose
  - Langzeitarbeitslose
  - Transferbezieher
  - Transferbezieher unter 15 Jahre
- Weitere Kontextindikatoren wie Wanderungssaldo



#### **OPENSTREETMAP**

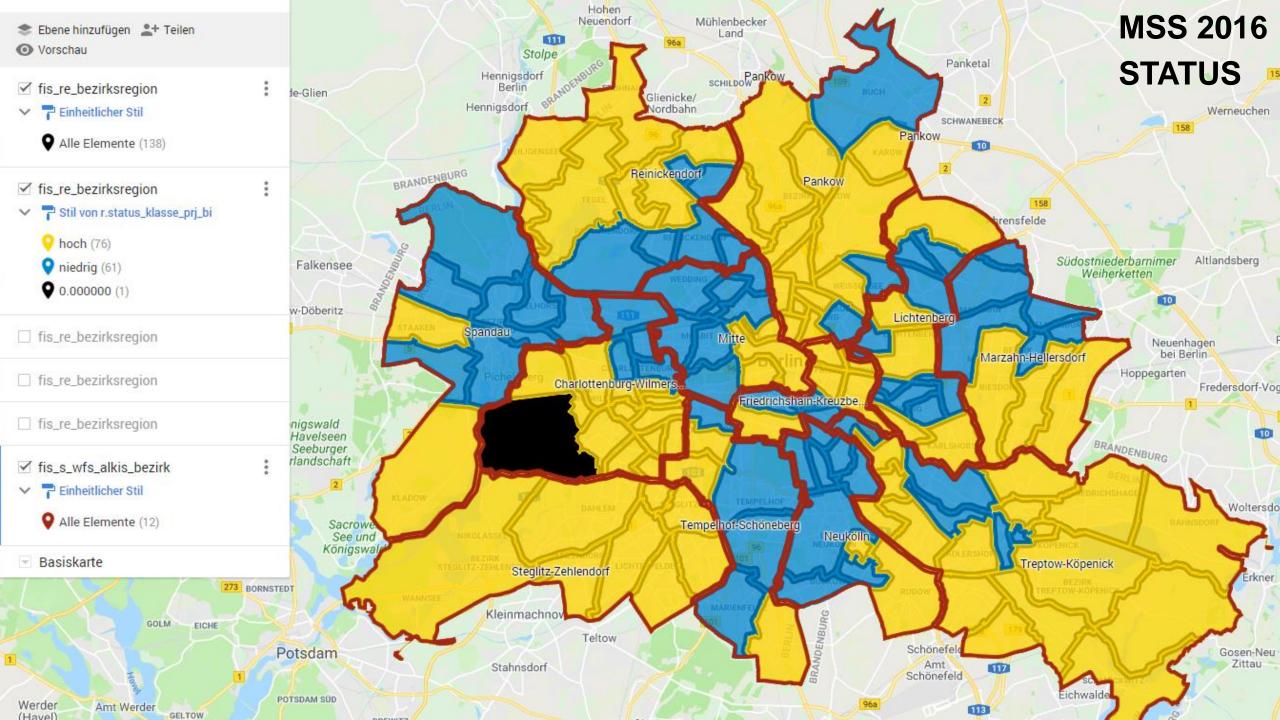
- Ab 01.2014 jährlich verfügbar bei Geofabrik GmbH
- OSM hat nur drei Datentypen
  - Node, Way, Relation
- Alle weiteren Informationen über KEY-VALUE Liste
  - Mapper-Community nutzt standardisierte Kombinationen

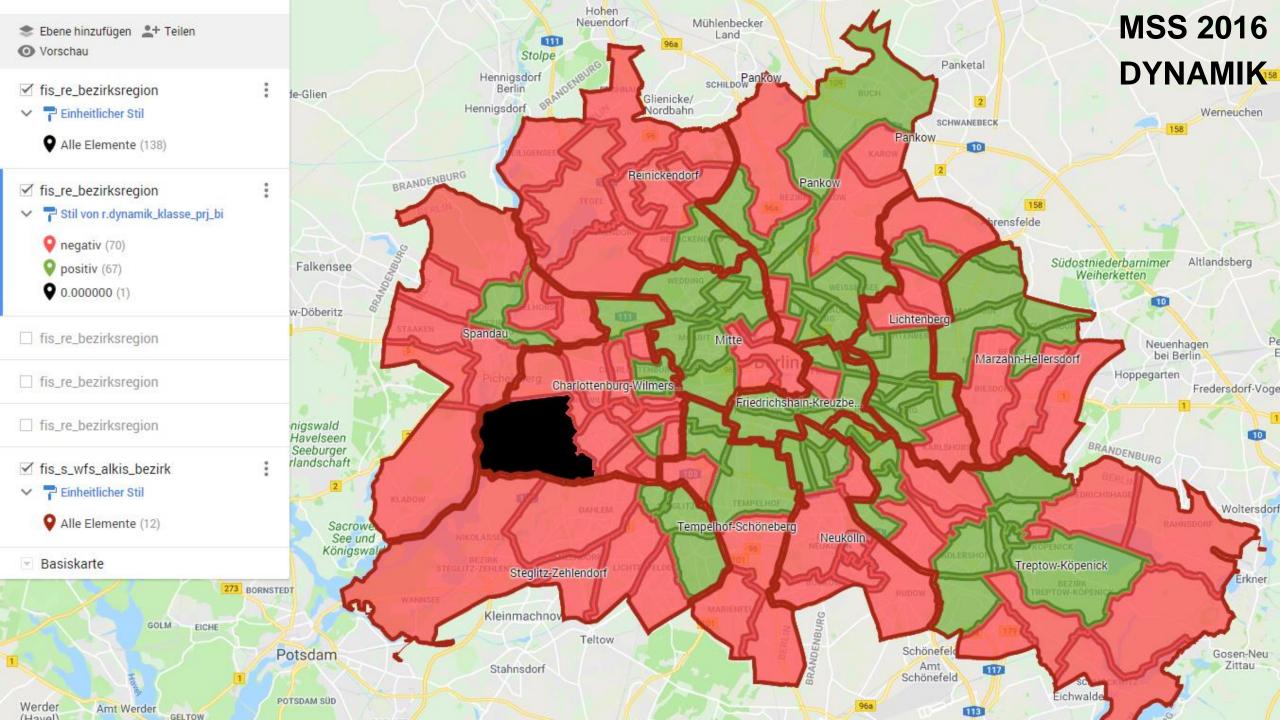
<tag k="amenity" v="restaurant"/>
<tag k="cuisine" v="italian"/>
<tag k="name" v="Hannibal"/>
<tag k="opening\_hours" v="Su-Th 09:00-02:00, Fr,Sa 09:00-04:00"/>



- Eigens entwickeltes Mapping auf OSM KeyValue Tags
- Abbildung als Hierarchie für Aggregationen

Domain	Category	Туре
Gastronomie	Restaurant	Restaurant Italiener
Gastronomie	Restaurant	Restaurant Italiener
Dienstleistun g	Friseur	Friseur
Waren	Essen & Trinken	Baeckerei
Waren	Kleidung	Kleidung





### Die technologische Grundlage für die Datenintegration bildete ein Hadoop-Cluster, basierend auf Hortonworks HDP. Die Daten wurden in einem ETL-Prozess aufbereitet.

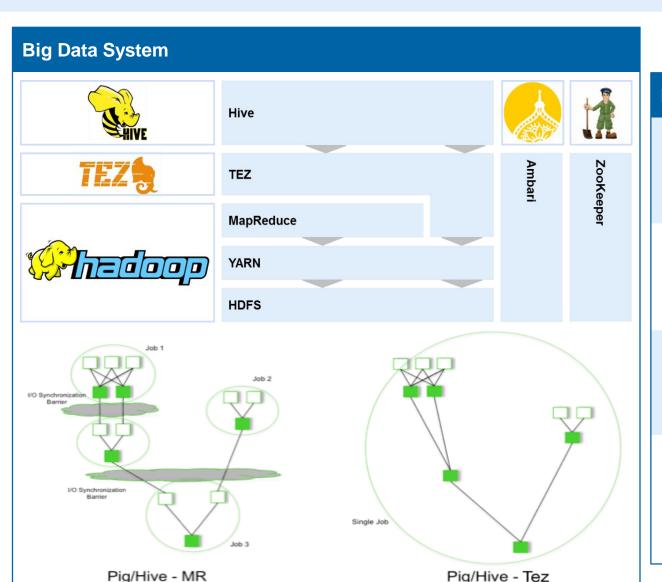


**Data Preparation – System und ETL-Prozess** 









#### **ETL-Prozess**

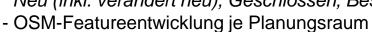
### POI-Daten aus OpenStreetMap erzeugen

- Relevante Nodes aus OSM extrahieren
  - Planungsraum auf Basis Koordinaten ermitteln
  - POI-Typ anhand von Key-Value Feldern mappen

### **OSM-POI-Daten in Raumdaten aggregieren**

- POI-Statuserkennung im Zeitverlauf:

Neu (inkl. verändert neu), Geschlossen, Bestand



### Öffentliche Datenquellen anbinden

- Indexaufbau nach Döring & Ulbricht (Dynamik)
  - MSS-Daten (Dynamik und Status)
  - Standardisierung (z-Transformation) & binäre Klassen

### Weitere (nachträgliche) Berechnungen

- Einwohnergewichtete POI-Kennzahlen
  - Distanzgewichtete Aggregation
  - Aggregation über LOR-Hierarchie
  - Berechnung Offering Advantage

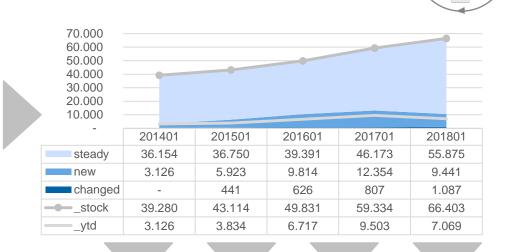


# Die Transformation der OSM-Daten hat den Großteil der Datentransformation ausgemacht. Im Ergebnis wurde für jede Hypothese ein Datensatz generiert.



### **Data Preparation – Ergebnis**

	41 11				
node_id	timeslice	planungsraum	poi_type	name	poi_state
N1000710165	201501	4030931	fast_food Sonstiges	China Box	steady
N1000710165	201601	4030931	fast_food Sonstiges	China Box	deleted
N1000710165	201701	4030931	null	null	null
N1552074537	201401	1011302	null	null	null
N1552074537	201501	1011302	Restaurant Steakhouse	Shiso Burger	new
N1552074537	201601	1011302	Restaurant Steakhouse	Shiso Burger	steady
N26735763	201701	4030828	Restaurant Sushi	Sakana	steady
N26735763	201801	4030828	Restaurant Asiatisch	Lemongrass	changed
N26735763	201801	4030828	Restaurant Sushi	Sakana	changed_deleted



Hypothese		Gent	rifizierung Sozial I	Indexwert	kwert POI Featuregruppen			
		MSS-Status	MSS-Dynamik	Döring & Ulbricht	OSM_stock OA	OSM_stock	OSM_new	OSM_ytd
H1 (a-c)	IST – IST	2016_EOY			2016_E	EOY		
H2	IST – YTD (SozialDyn durch POI-IST)		2014_EOY - 2016_EOY		2014_EOY			
Н3а	YTD – YTD (SozialDyn nach POI)		2014_EO	Y – 2016_EOY			2012_EOY	– 2014_EOY
H3b	YTD – YTD (SozialDyn vor POI)		2012_EO\	Y – 2014_EOY			2014_EOY	– 2016_EOY
Н3с	YTD – YTD (SozialDyn gleichz. POI)		2014_EOY - 2016_EOY				2014_EOY	- 2016_EOY
			Anzahl Dat	enfeatures (Spalten)	85	229	229	229

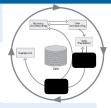
# Auf Basis der Ergebnisse der Datenintegration wurden mit Weka Modelle gebildet und gleichzeitig evaluiert. Im ersten Schritt wurde die beste Datenbasis evaluiert.



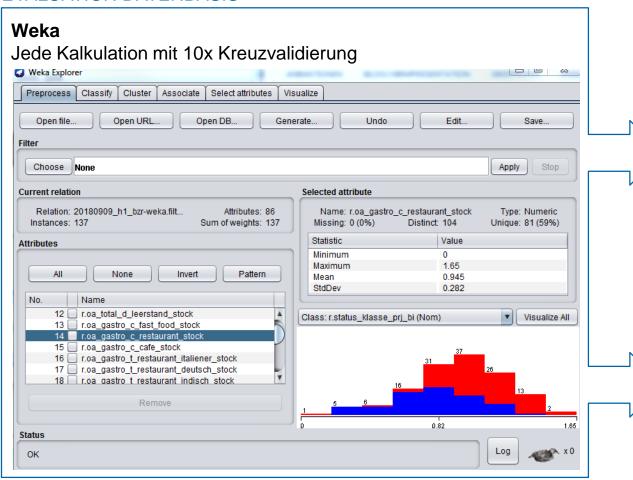
**Modeling & Evaluation** 



3



#### **EVALUATION DATENBASIS**



### Raumbezugsgröße

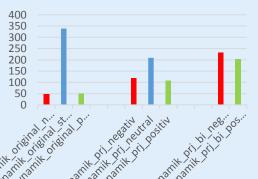
- Distanzgewichtung der Planungsräume verbessert Modelle auch
- "Herauszoomen" in Bezirksraum verbessert Modelle deutlich
- Reine PLR-Betrachtung liefert schlechte Modelle
- → Bezirksraum + Distanzgewichtet\_Planungsraum

### Featuregruppen

- Offering Advantage verbessert Modelle deutlich  $\frac{count(Restaurant_{Italiener}, Wrangelkiez)}{count(Gastronomie, Wrangelkiez)} \cdot \frac{count(Gastronomie)}{count(Restaurant_{Italiener})}$
- Nur NEU-Kennzahl besser als NEU und YTD
- → Nur NEU

#### Indexwerte

- Binäre Indexwerte ergeben bessere Modelle
- Dynamik mit MSS besser als komplexe DöringUlbricht-Kennzahl
- → MSS + binär (Status & Dynamik)



# Aufbauend auf der Datenbasis wurden im folgenden 16 Algorithmen parametrisiert und mit 10 Datensätzen ausgeführt (5 Hypothesen Datensätze x 2 Raumbezugsgrößen)



**Modeling & Evaluation** 







#### **EVALUATION ALGORITHMEN**

Algorithmus	Anzahl	Parameter
RandomForest	3	Std / max. BTiefe = 5 / max. BTiefe = 3
LMT	2	Std / fixe Anzahl LogitBoost-Iteration = 3
LogitBoost	1	Std
SimpleLogistic	2	Std / fixe Anzahl LogitBoost-Iteration = 3
BayesNet	1	Std
NaiveBayes	1	Std
J48	1	Std
AdaBoostM1	1	Algorithmus = J48
Stacking	1	Meta-Algo: J48 Algo: 3xRF, 2xLMT, 2xSL (s.o.)
Bagging	1	Algo: SL Std
Voting	1	Algo: 3xRF, 2xLMT, 2xSL (s.o.)
MultilayerPerceptron	1	20 Layer

	Algorithmus	AUC	F1
	02 - RandomForest	<b>0</b> ,74	<b>0</b> ,67
	03 - RandomForest_5	<b>0</b> ,73	<b>0</b> ,67
	04 - RandomForest_3	0,72	• 0,66
	05 - LMT	0,68	0,64
>	06 - LMT_3	0,69	0,64
	08 - LogitBoost	<b>0</b> ,69	0,64
	09 - SimpleLogistic	0,67	• 0,63
	10 - SimpleLogistic_3	0,68	0,62
	11 - BayesNet	0,68	0,63
	12 - NaiveBayes	0,61	0,54
	13 - J48	0,59	0,58
	14 - AdaBoostM1	0,65	0,61
>	15 - Stacking	0,65	0,65
	16 - Bagging	• 0,70	0,63
	18 - Vote	0,73	• 0,66
	19 - MultilayerPerceptron	0,64	• 0,58

- RandomForest liefert fast immer die besten Modelle
- Modelle mit logistischer Regression schneiden durch die vielen Features gut ab
- NaiveBayes wegen vieler Features vorhersehbar schlecht
- Ensembles wegen enthaltender RFs und LogReg ebenfalls relativ gut

Im Ergebnis wurden alle Hypothesen bewertet. H1 belegt einen starken Zusammenhang zwischen lokaler Angebotsstr. und sozialem Status. H3a (Soziale Veränderung nach Änderung POI) konnte verworfen werden, das Gegenteil (H3b – VOR) dagegen eher bestätigt.

Modeling & Evaluation

Hypothese		Bestes Mach	Bestes Machine Learning Model			Korrelations	Bewertung	
ID	Bedeutung	Algorithmus	AUC	F1		Anzahl Features mit cor > 0,3	Features sinnvoll?	
H1	IST- IST (Sozial – POI)	SimpleLogistic_3	0,87	0,79	ja	8	ja	✓
H1a	-cafe positiv							0
H1b	-fastfood negativ							✓
H1c	-sport positiv							0
H2	IST – YTD (SozialDyn durch POI-IST)	RandomForest	0,77	0,65	eher nein	1	ja	0
НЗа	YTD – YTD (SozialDyn nach POI)	LogiBoost	0,72	0,67	nein	0	-	X
H3b	YTD – YTD (SozialDyn vor POI)	Bagging _SimpleLogistic	0,81	0,72	eher ja	3	ja	(√)
Н3с	YTD – YTD (SozialDvn gleichz, POI)	LogiBoost_2	0,71	0,77	ja	1	nein	0

### Class niedrig: 0.9

+ -1.53 \* [r.oa\_gastro\_c\_restaurant\_st

+ 0.3 \* [r.oa\_pubserv\_c\_sozial\_stock]

+ 0.15 \* [r.oa\_vergnuegung\_t\_spieloth

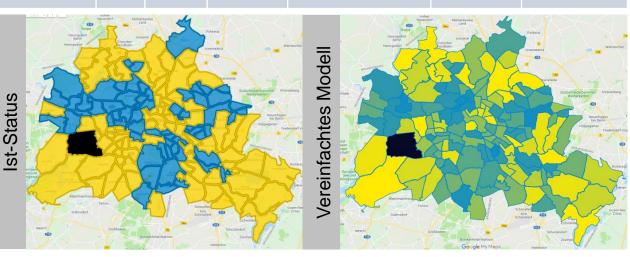
#### Class hoch:

-0.9

+ 1.53 \* [r.oa\_gastro\_c\_restaurant\_sto

+ -0.3 \* [r.oa pubserv c sozial stock

+ -0.15 \* [r.oa vergnuegung t spielot



### H1/BZR

cor	Feature	Entarties
-0,45	r.oa_gastro_c_restaura	int_stock
0,44	r.oa_gastro_c_fast_foo	d_stock
-0,43	r.t_briefkasten_stock	
-0,43	r.c_briefe_und_pakete_	_stock
0,41	r.oa_vergnuegung_t_sp	oielothek_stock
-0,41	r.t_haltestelle_stock	
-0,40	r.c_oepnv_stock	
0,39	r.t_spielothek_stock	
-0,37	r.oa_gastro_t_restaura	nt_deutsch_sto
0,35	r.c_zwielicht_stock	
0,32	r.oa_total_d_vergnuegu	ung_stock

UHI

### H3b / BZR

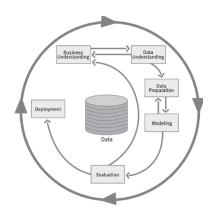
or	Feature
-0,37	r.t_bar_new
-0,36	r.c_gaststaetten_new
-0,34	r.d_vergnuegung_new
-0,32	r.t_fahrrad_new
0,31	r.t_haltestelle_new
-0,30	r.t_bar_ytd
-0,29	r.t_pub_new



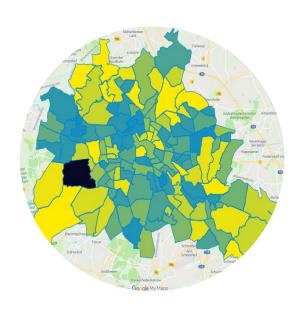




Motivation, Ziel der Arbeit und Forschungsfragen



Theoretische Grundlagen und methodisches Vorgehen



Diskussion der Ergebnisse und Ausblick

# Die Masterarbeit unterlag durch die Auswahl der Datenquellen einigen Limitierungen. Im Ergebnis konnten jedoch relevante Erkenntnisse gewonnen werden.



### **Diskussion**

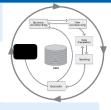
#### LIMITIERUNGEN

- Immobilienwirtschaftliche Aufwertung
  - Keine Daten angebunden
- Verdrängung und soziale Aufwertung
  - Nur die schwächsten enthalten, keine Daten zum Mittelstand
- Veränderung der Angebotsstruktur
  - OSM relativ wenig POI-Datensätze
  - OSM-Datenqualität

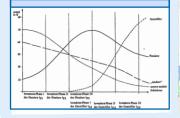
#### Zeitverlauf der POI-Kennzahlen in Berlin



#### **ERKENNTNISSE**



- Starker Zusammenhang zwischen lokaler Angebotsstruktur und sozialem Status auf BZR-Ebene
  - Restaurants in Nachbarschaften mit höherem Status
  - Fast-Food in Nachbarschaften mit niedrigem Status
- Änderung der lokalen Angebotsstruktur folgt vermutlich zeitlich auf Änderung des sozialen Status
  - Anzeichen für die erste Phase des doppelten Invasions-Sukzessions-Zyklus nach Dangschat:
    - sozial schwache Bevölkerung von Pionieren verdrängt
    - Pioniere bringen Kultur und Angebotsstruktur
  - Verdrängung Pioniere durch Gentrifizierer folgt?!
    - → Qualitative Folgeuntersuchung!!





### Frage:

### "Würde eine gute Prognose die Gentrifizierung eher beschränken, oder sogar noch weiter beschleunigen?"



**Stellungnahme** 



### WELCHE MAßNAHMEN KÖNNTE DIE POLITIK ZIEHEN?

- Die Politik kann nicht entscheiden, wo Menschen wohnen sollen
- Die Politik kann nicht verhindern, dass Menschen Wohnraum verkaufen
- Die Politik kann Preise nicht diktieren
- ABER: Politik kann Vorkaufsrecht nutzen (siehe Hamburg)

2

### SELBSTERFÜLLENDE PROPHEZEIUNG

- Frühzeitiger Zuzug von Menschen, die sich "hippe" Nachbarschaft wünschen
- Menschen könnten versuchen "günstige" Mietverträge zu bekommen
  - Teurer untervermieten
  - Airbnb
  - Selbst dort wohnen



### SOLANGE BILLIGES GELD DA IST, BLEIBT IMMO-BOOM

- Rasanter Anstieg der Immobilienpreise in Berlin:
   21% in einem Jahr
- Bei niedrigen Zinsen weitere Investitionen wahrscheinlich
- Wirtschaft würde Modelle schnell adaptieren und Gewinne maximieren
- Bei Banken dennoch keine Sicherheit

  A Nur Immobilionfonds
  - → Nur Immobilienfonds können in teure Immobilien investieren



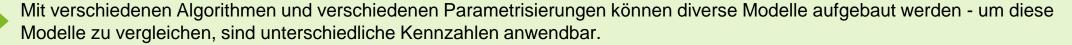
- Politik begrenzter Spielraum
- Bürger könnten Modell für Mietentscheidung nutzen
- Billiges Geld führt weiterhin zum Immobilienboom

# Backup

### Je nach verwendeter Methode sowie Untersuchungsfokus bieten sich unterschiedliche Performancemaße an – in der Masterarbeit wurden F1 und AUC verwendet.



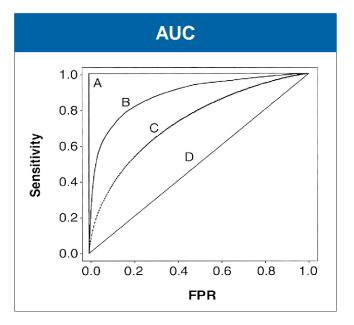
### Übersicht Performancemessung





Vorhe	Vorhersage		Vorhe	rsage			
Α	В			A	В		
5	5	Α	lst-\	TP	FN	Α	Ist-V
2	8	В	Ist-Wert	FP	TN	В	lst-Wert
F-Measure	2 * Reca	II * Precision /	(Recall + P	recision)		chnung des F- s des Modells	
Recall (Sensitivity)		TP / (TP	+ FN)		Mittel der	harmonische F-Measures	
Precision		TP / (TP	+ FP)		über alle gebildet.	Klassen	

FP/(FP + TN)



**FP Rate** 



### **Modeling & Evaluation**

### **EVALUATION ALGORITHMEN - DETAIL**

Measure Dataset [2] [3] [4] [5] [6] [8] [9] [10] [11] [12] [13] [14] [15] [16] [18] [19] MAX Bewertung MEAN Rewertung Weight-awg AUC hl_plin-dc Wei																			•			
Weight.avg AUC h1_plr-dc Weight.avg AUC h2_brr	Measure	Dataset	[2]	[3]	[4]	[5]	[6]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]					MEAN	Bewertung
Weight.avg AUC h2_plr-dc	Weight.avg AUC	h1_bzr	0,86	0,85	0,84	• 0,86	0,86	0,78	0,86	0,87	0,82	0,70	0,63	0,76	0,76	0,82	0,88	0,73	0,88	AUC - B	0,80	AUC - B
Weight.avg AUC h2plr-dc weight.avg AUC h3a_bzr weight.avg AUC h3b_bzr weight.avg AUC h3b_bz	Weight.avg AUC	h1_plr-dc	0,82	0,82	• 0,80	• 0,80	• 0,77	0,79	• 0,80	0,72	0,77	0,70	0,65	0,75	0,73	0,79	0,83	0,76	0,83	AUC - B	0,76	AUC - C
Weight.avg AUC h3a_bzr	Weight.avg AUC	h2_bzr	0,77	0,73	0,71	0,62	0,61	0,60	0,60	0,63	0,56	0,59	0,47	0,56	0,57	0,70	0,67	0,67	0,77	AUC - C	0,62	AUC - D
Weight.avg AUC h3a_plr-dc weight.avg AUC h3b_bzr	Weight.avg AUC	h2_plr-dc	0,73	0,72	0,73	0,66	0,70	0,66	0,66	0,70	0,68	0,66	0,58	0,65	0,67	0,68	0,72	0,65	0,73	AUC - C	0,67	AUC - D
Weight.avg AUC h3a_plr-dc	Weight.avg AUC	h3a_bzr	0,53	0,57	0,57	0,55	0,60	0,56	0,55	0,61	0,50	0,54	0,51	0,47	0,53	0,51	0,58	0,45	0,61	AUC - D	0,54	AUC - F
Weight.avg AUC   h3b_bzr   weight.avg AUC   h3b_bzr   weight.avg AUC   h3b_plr-dc   h3b_plr-dc   weight.avg AUC   h3b_plr-dc   h3b_plr-dc   weight.avg AUC   h3b_plr-dc   h3			0,70	0,68	0,67	0,57	0,56	0,72	0,55	0,50	0,65	0,53	0,58	0,66	0,64	0,62	0,64	0,61	0,72	AUC - C	0,61	AUC - D
Weight.avg AUC   N3b_plr-dc   0,74   0,72   0,65   0,65   0,65   0,66   0,64   0,67   0,70   0,59   0,61   0,67   0,64   0,67   0,71   0,58   0,74   AUC - C   0,66   AUC - D   Weight.avg AUC   N3c_plr-dc   0,75   0,74   0,72   0,70   0,75   0,74   0,69   0,75   0,74   0,69   0,71   0,74   0,59   0,66   0,66   0,67   0,68   0,60   0,68   0,60   0,68   0,60   0,68   0,70   0,75   0,74   0,75   0,74   0,75   0,74   0,75   0,74   0,75   0,74   0,75   0,74   0,75   0,75   0,74   0,75   0,75   0,74   0,75   0,75   0,74   0,75   0,75   0,74   0,75																						
Weight.avg AUC h3c_brr h3c_br h3c_brr h3c_br h3c_brr h3c_brr h3c_brr h3c_brr h3c_brr h3c_brr h3c_brr h																		_			_	
Weight.avg F1																	_					
Weight.avg F1 h1_bzr																						
Weight.avg F1 h1_bzr	3 3						_														,	
Weight.avg F1 h1_plr-dc	Weight.avg F1																			F1 - B	0,73	F1 - B
Weight.avg F1  h2_ptr-dc																					_	
Weight.avg F1 h2_plr-dc																					0,58	F1 - D
Weight.avg F1 h3a_bzr																						
Weight.avg F1 h3a_plr-dc		_																				
Weight.avg F1 h3b_bzr																						
Weight.avg F1 h3b_plr-dc		_																				
Weight.avg F1 h3c_bzr																						
Weight.avg F1 h3c_plr-dc 0,67 0,67 0,69 0,58 0,65 0,65 0,65 0,59 0,59 0,59 0,59 0,67 0,42 0,63 0,62 0,67 0,65 0,65 0,65 0,52 0,70 F1 - B																						
									_									_			_	
																					0,01	

Algorithmus
02 - RandomForest
03 - RandomForest_5
04 - RandomForest_3
05 - LMT
06 - LMT_3
08 - LogitBoost
09 - SimpleLogistic
10 - SimpleLogistic_3
11 - BayesNet
12 - NaiveBayes
13 - J48
14 - AdaBoostM1
15 - Stacking
16 - Bagging
18 - Vote
19 - MultilayerPerceptron