

Vorhersage des Privatbesitzes von Straßen mit Hilfe von Machine Learning

CHRISTOPH RÖHRL

GUTACHTERIN: PROF. DR.
HANNAH BAST

BETREUER: DR. BRUNO BECKER,
DIPL.-INF. MICHAEL SCHLENK

ALBERT-LUDWIGS-UNIVERSITÄT
FREIBURG
TECHNISCHE FAKULTÄT
INSTITUT FÜR INFORMATIK
PROFESSUR FÜR
ALGORITHMEN UND
DATENSTRUKTUREN

12. JULI 2023



Übersicht

1. Einleitung

2. Umsetzung

- Datenquellen
- Bereinigung/Verarbeitung der Daten
- Trainieren von zwei verschiedenen ML-Modellen

3. Evaluation

- Leistungsvergleich
- Aussagekraft versch. Attribute
- Ausblick

Einleitung

- Deutschland ist beim Glasfaserausbau noch in frühen Entwicklungsstadium (7,1 %).
- Der Bedarf wächst immer mehr an.
- Dem Anschluss und Ausbau von Glasfaser geht aufwendiger Planungsprozess vorher.
- Ein Faktor, welcher Kosten und Komplexität eines Ausbauprojektes signifikant beeinflusst ist der Privatbesitz von Straßen.
- Ausbau in Privaten Straßen verursacht einen höheren Bürokratie- und Planungsaufwand.
- Um die dadurch entstehenden Kosten gering zu halten, sollten solche Straßen vermieden werden.

Einleitung II

- Viele private Straßen lassen sich aufgrund ihrer öffentlichen Widmung nicht als solche erkennen.
- Das Ermitteln der Besitzumsverhältnisse von Straßen durch Kontakt mit Behörden ist aufwendig und Zeitintensiv.
- In dieser Arbeit wurde untersucht, inwiefern sich der Privatbesitz von Straßen mithilfe von öffentlich zugänglichen OSM-Daten und Machine Learning vorhersagen lässt.

Einleitung III

OSM Straßen und deren Attribute
(Geometrie, Länge, Höchstgeschwindigkeit, Oberfläche, ...)



Datenverarbeitung, Machine Learning



Vorhersage Privatbesitz: Ja/Nein

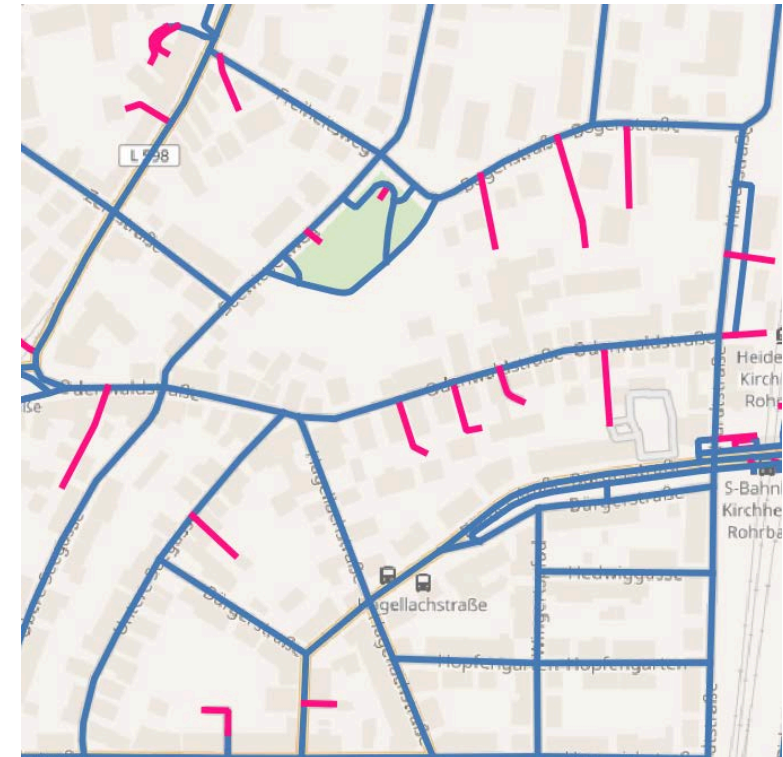
Umsetzung

Umsetzung - Datenquellen

- OpenStreetMap:
 - Internationales Open-Source Projekt, welches Weltweit Kartendaten lizenzkostenfrei zu verfügung stellt
 - Daten werden von freiwilligen Helfer:innen gesammelt
 - Detailinformationen über Straßen und andere Objekte liegen als ‘Tags’ vor
 - Qualität & Abdeckung der Tags unterliegt starken schwankungen
- Amtliche Daten der Stadt Heidelberg:
 - Zur verfügung gestellt vom Tiefbauamt Heidelberg
 - Shape Datei welche alle Straßenverkehrsflächen im Besitz der Stadt beinhaltet

Umsetzung – Verarbeitung der Daten

- Erstellen einer Grundwarheit
 - Ausschneiden aller Städtischen Straßen aus den OSM Straßen mittels QGIS
 - Behandlung von Randfällen: Forstwege, Autobahnen, Cut-Off
- Berechnen von Sackgassen
 - Herausfiltern von alleinstehenden “Locations” der Geometrien
 - Wenn Nebenbedingungen erfüllt sind stellen diese eine Sackgasse dar
- Bereinigen der Daten
 - Entfernen von nicht-relevanten Straßentypen
 - Entfernen/Ersetzen von unbrauchbaren Tag Werten



Umsetzung – Trainieren von ML-Modellen

- Neuronales Netzwerk
 - Umgesetzt mit dem *PyTorch Tabular* Framework
 - Aufteilung in 90% Trainingsdaten und 10% Testdaten durch Pseudozufall
 - Training dauert nur wenige Minuten

- Gradient Boost Modell
 - Umgesetzt mit dem *XGBoost* Framework
 - Gleiche Aufteilung der Daten wie beim NN
 - One-Hot Codierung der Daten
 - Training dauert ebenfalls nur wenige Minuten



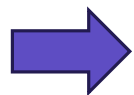
Evaluation

Evaluation - Leistungsvergleich

Metrik	Neuronales Netzwerk	Gradient Boost Modell
Accuracy	0.82	0.78
Precision	0.86	0.78
Recall	0.85	0.88

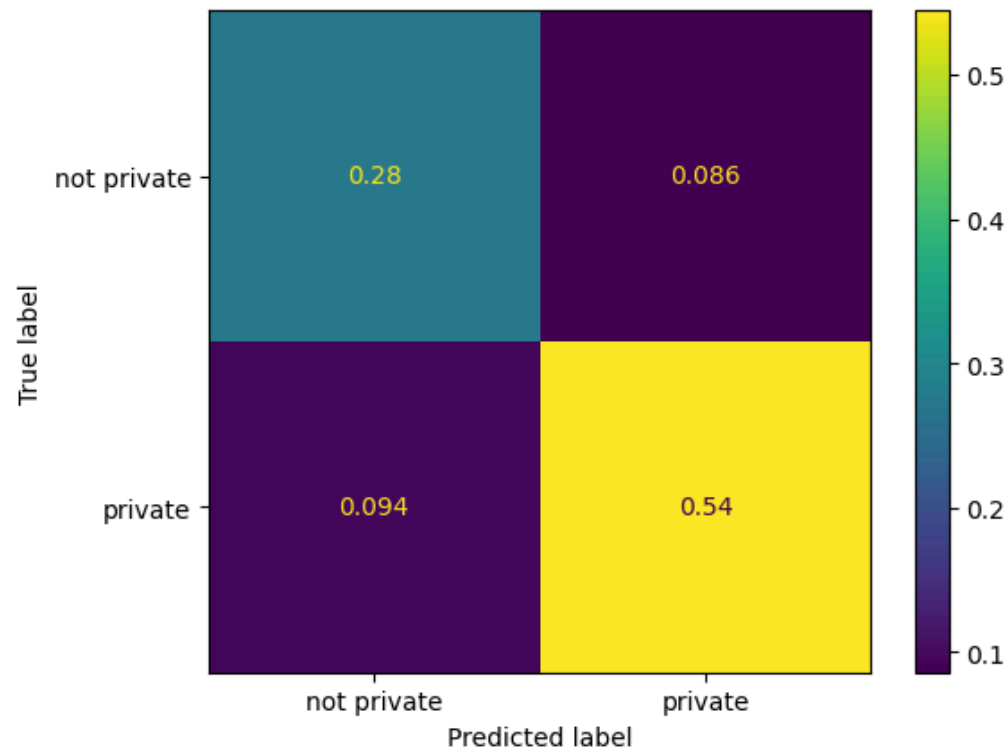
Ergebnisse des Trainings beider Methoden im Vergleich

Grundwarheit: 64% true, 36% false

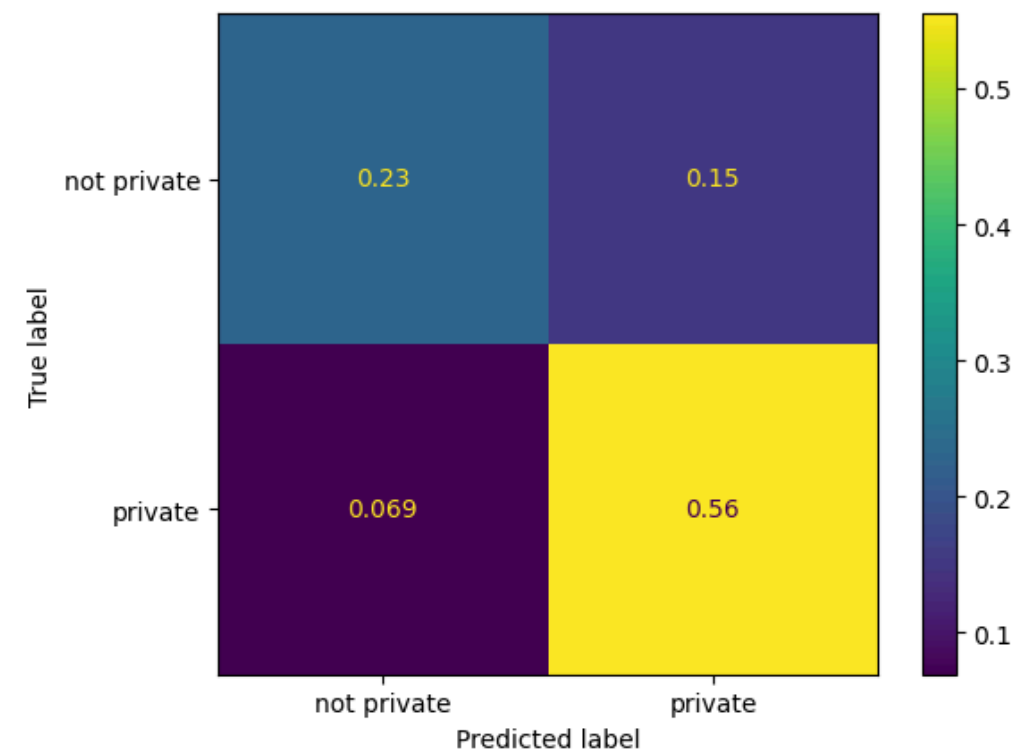


Beide Modelle haben zumindest bis zu einem gewissen Grad komplexere Zusammenhänge gelernt.

Evaluation – Leistungsvergleich II



Neuronales Netzwerk

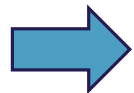


Gradient Boost Modell

Evaluation – Aussagekraft der Attribute

Neuronales Netzwerk	Gradient Boost Modell
type (-0.194)	type_service (-0.129)
service (-0.013)	access_unknown (-0.015)
access(-0.011)	maxspeed (-0.015)

Top drei Permutation Feature Importance Scores beider Modelle



Vor allem der Straßentyp ‘Service’ scheint in beiden Modellen eine wichtige Rolle zu spielen.

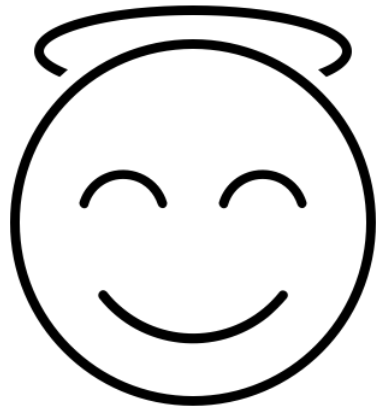
Evaluation – Aussagekraft der Attribute II

Attribut	Korrelationskoeffizient
type_residential	-0.49
surface_asphalt	-0.38
service_no	-0.34
.	.
.	.
simple_deadend	0.22
sidewalk_none	0.35
max_speed	0.38
type_service	0.59

Korrelationskoeffizienten der Grundwahrheit mit allen anderen Attributen

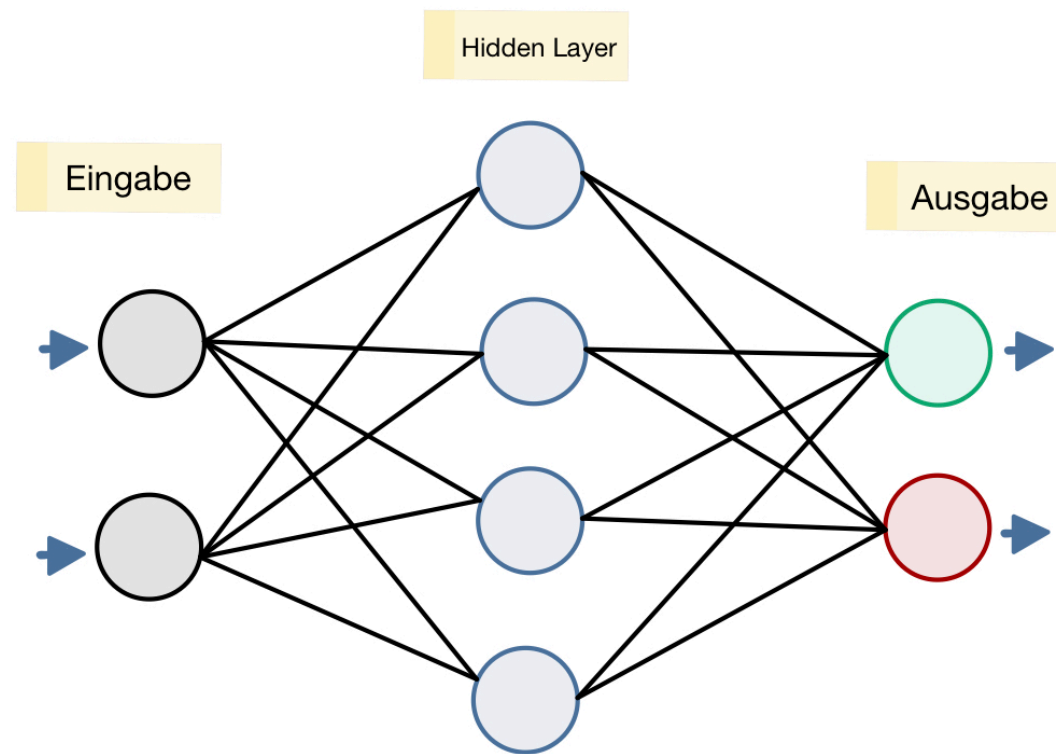
Evaluation - Ausblick

- Die Evaluation hat gezeigt, dass das Vorhersagen des Privatbesitzes von Straßen mittels OSM-Daten grundsätzlich möglich ist.
- Um die Genauigkeit und Robustheit zu erhöhen, könnten einige Verbesserungen vorgenommen werden:
 - Benutzen einer Größeren Datenmenge für das Training
 - Straßen aus mehreren Städten verwenden (nicht nur Heidelberg)
 - Neben Sackgassen auch noch weiteres “U-Form” Attribut berechnen
- Um die hier getesteten Methoden in der Praxis anzuwenden, müssten zudem Anpassungen im Hinblick auf die Quote der False-Positives vorgenommen werden.



Danke für Ihre
Aufmerksamkeit!

Backup I



Backup II

