

OPENDATA KI

meets



ABSCHLUSSWORKSHOP 23.03.2018

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

DAS PROJEKT ODAKI

- Open-Data-Innovationen mit KI-Methoden
- Untersuchung der Möglichkeiten und Vorteile von Methoden der Künstlichen Intelligenz (KI) bei der Entwicklung innovativer Open-Data- Anwendungen
- Laufzeit: 1.8.2017 – 31.3.2018

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

DAS ODAKI-TEAM

HITeC
HAMBURGER INFORMATIK
TECHNOLOGIE-CENTER E.V. (HITeC)

in Kooperation mit

GOVDATA
Das Datenportal für Deutschland
GESCHÄFTS- UND
KOORDINIERUNGSGESTELLE (GKST)



Stephanie von Riegen



Prof. Dr. Bernd Neumann



Christian Horn



Dr. Lothar Hotz



Yen Dieu Pham

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

AGENDA

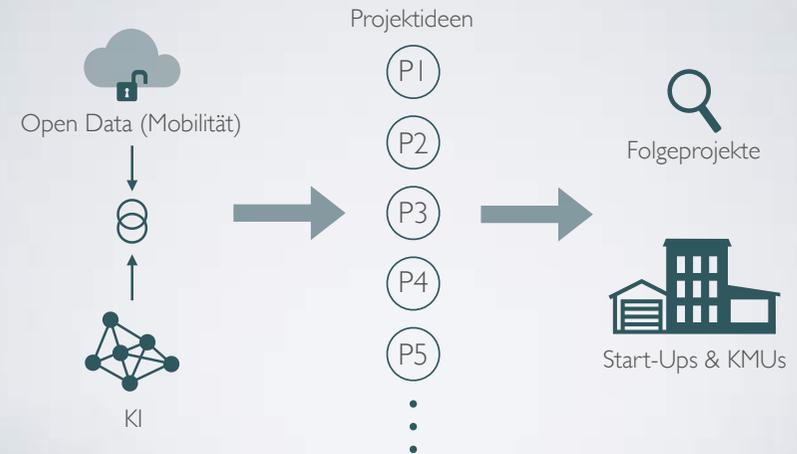
- 1 Projektrahmen
- 2 KI-Methoden
- 3 Ideenskizzen & Vorstudien
- 4 Erfahrungen & Ausblick
- 5 Diskussion

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



PROJEKTRAHMEN

ZIEL



Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

TEILNAHME AN...

- ... International Joint Conference on Artificial Intelligence, Melbourne
- ... Co-Creation-Workshop Berlin, INNOZ
- ... KI-Konferenz, Dortmund
- ... Open Data Science Conference, London
- ... ITS Hackathon DB, Hamburg
- ... Mobility Hackathon Hochbahn, Hamburg
- ... Data Run Hackathon BMVI, Berlin
- ... m-Fund Begleitforschung (Wissenschaftliches Institut für Infrastruktur & Kommunikationsdienste Bonn)



Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

LEITGEDANKEN OPEN DATA

- Förderung von **Transparenz** und **Zusammenarbeit**
- **Nutzbarmachung** der Daten, nicht nur der **öffentlichen Verwaltung**, sondern auch der **Privatwirtschaft (open corporate data)**
 - Schätzungen des volkswirtschaftliche Potenzials auf 30Mrd €/Jahr und etwa 20.000 neue Arbeitsplätze
- Datennutzung/-verbreitung ohne jedwede Einschränkung
- **Voraussetzungen** für die Datensätze:
 - **maschinenlesbares und standardisiertes** Format
 - **rechtliche Offenheit**

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

PLATTFORMEN & DATENSÄTZE



Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

15 INTERVIEWS



VIELEN DANK FÜR DIE ZEIT !

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

INTERVIEW-ERGEBNISSE (1)

- Großes Interesse an der KI und einer KI-gestützten Anwendung
- Probleme und Vorbehalte bei der Bereitstellung von Datensätzen als Open Data
- Daten sind wirtschaftlich relevant
- Wunsch nach kontinuierlichen, automatisierten Open-Data-Datensätzen
- Bedarf an einem akkreditierten Anonymisierungsverfahren für Daten-Provider
- Interesse an einem Simulator für Planungsaufgaben
 - Verkehr: Rad-, Fußgänger-, ÖPN-, ÖPF-, Automobilverkehr
 - Einbeziehen von Einflussgrößen: Ferien, Wetter, Veranstaltungen, Bebauung...

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

INTERVIEW-ERGEBNISSE (2)

- Kennenlernen von Kunden und Kundenverhalten
- Verkehrsverhalten einzelner Verkehrsteilnehmer (von Tür zu Tür)
- Dazu häufige Wünsche:
 - Prognoseerstellung,
 - Optimierungen (bei Planung und Wartung)
 - Erfassung von Verkehrsstärken
 - Entscheidungshilfen
 - Erforderliche Daten stehen nicht immer als Open Data zur Verfügung

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

2

KI - METHODEN

KI IST NICHT NEU...

- 1950: Alan Turing "Computing Machinery and Intelligence"
- 1957: Arbeitstitel "Artificial Intelligence" entsteht in Dartmouth/USA
- 1967: Erstes Fachbuch Marvin Minsky "Computers and Thought"
- 1974: Fachgruppe "Mustererkennung und Künstliche Intelligenz" in der GI

... SEIT LANGEM ETABLIERTES FORSCHUNGSGEBIET

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

KI IST EIN UMFANGREICHES GEBIET

- Universitätskurse über mehrere Semester
- Beliebte Fachbücher:



(1200 Seiten)



(700 Seiten)

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



NEUE HERAUSFORDERUNGEN DURCH BIG DATA, WIE Z.B.:



Internet, soziale
Netzwerke



Digitalisierung



Internet-of-
Things (IoT)



Open Data,
Transparency

ABSCHLUSSBERICHT VON ODAKI

1. Zielsetzung.....	2
2. KI-Methoden zur Datenrepräsentation	3
2.1 OWL.....	4
2.2 Linked Open Data.....	5
2.3 Knowledge Graphs.....	6
2.4 Regelbasierte Systeme.....	6
2.5 Deduktive Datenbanken.....	6
2.6 Ontologieabgleich.....	7
3. KI-Methoden zur Datenanalyse und -aufbereitung	8
3.1 Suchverfahren.....	8
3.2 Lernverfahren.....	9
3.2.1 Überwachtes Lernen	9
3.2.2 Unüberwachtes Lernen	10
3.3 Fallbasiertes Schließen	11
3.4 Empfehlungsdienste	12
3.5 Bayes Netze.....	12
3.6 Planen	13
4. Bildverarbeitung	14
4.1 Objekterkennung.....	14
4.2 Szenenanalyse.....	14
4.3 Inhaltsbasierter Bildabruf.....	15
5. Sprachtechnologie.....	16
6. Robotik	16
7. Abschließende Bewertung.....	17

WICHTIGE KI-METHODEN FÜR ODAKI-THEMEN



Datenrepräsentation



Datenanalyse und
Datenaufbereitung



Bildverarbeitung



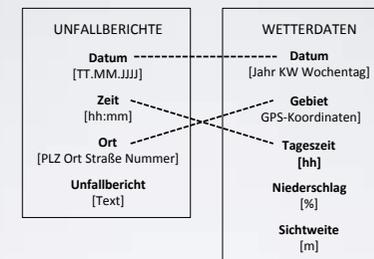
Sprachtechnologie



Robotik



INTEGRATION HETEROGENER DATEN



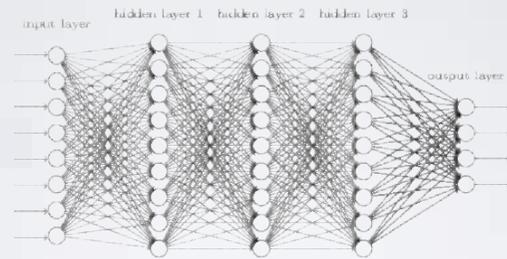
Strukturierte Wissensrepräsentation mit OWL (Web Ontology Language)

- bedeutungswahrende Integration durch gemeinsames Begriffssystem
- Schlussfolgerungen (Inferenzen) über implizite Zusammenhänge
- effiziente und teilweise standardisierte Tools



ANALYSE GROßER DATENBESTÄNDE: LERNVERFAHREN

NEURONALE NETZE (NN)



Überwachtes Lernen

Unüberwachtes Lernen

- **Nachteil: Fehlende Transparenz**

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

21.03.2018: <https://www.rspivision.com/exploring-deep-learning/>



ANALYSE GROßER DATENBESTÄNDE: BAYES NETZE



- Transparenz
- Vorhersagen
- Ursachenforschung

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



BILDVERARBEITUNG



- Fortschritte für Objekterkennung durch Deep Learning
- Szenenanalyse bisher vernachlässigt

Screenshot



SPRACHTECHNOLOGIE

- Klassische Analyse von Syntax und Semantik nur eingeschränkt erfolgreich
 - "Dem Fahrer seine Fahne war echt krass. Hatte wohl einen schweren Kater."
- Moderne KI wertet sprachlichen Kontext in umfangreichen Wissensnetzen aus
 - Auflösung von assoziativer Statistik
- Sentimentanalyse in sozialen Netzen
 - Erfolgreiche Vorhersage von Kinoerfolgen aufgrund von Vorschau-Blogs

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



ROBOTIK

"Robotik ist die intelligente Verbindung von Wahrnehmung und Handeln"

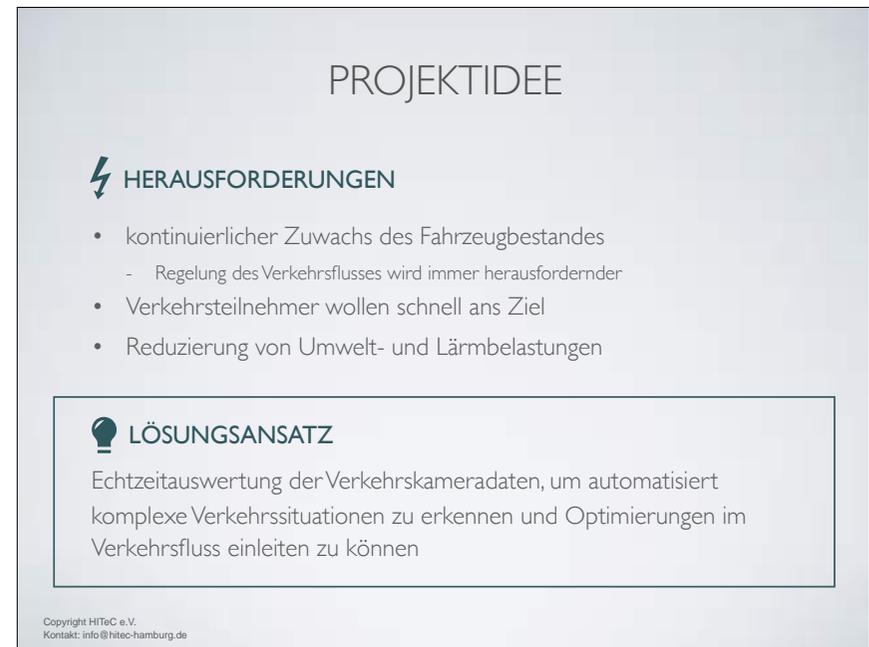
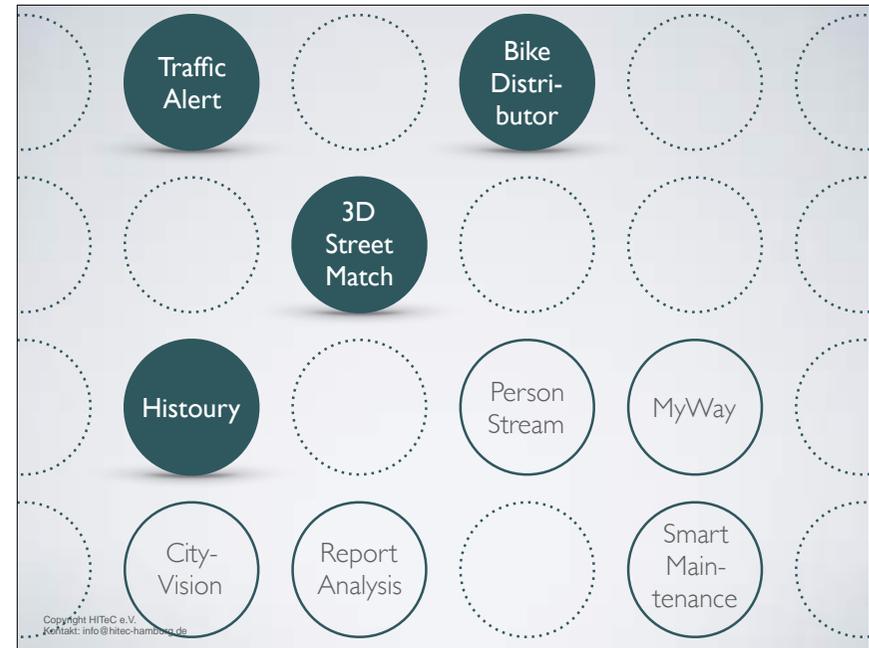
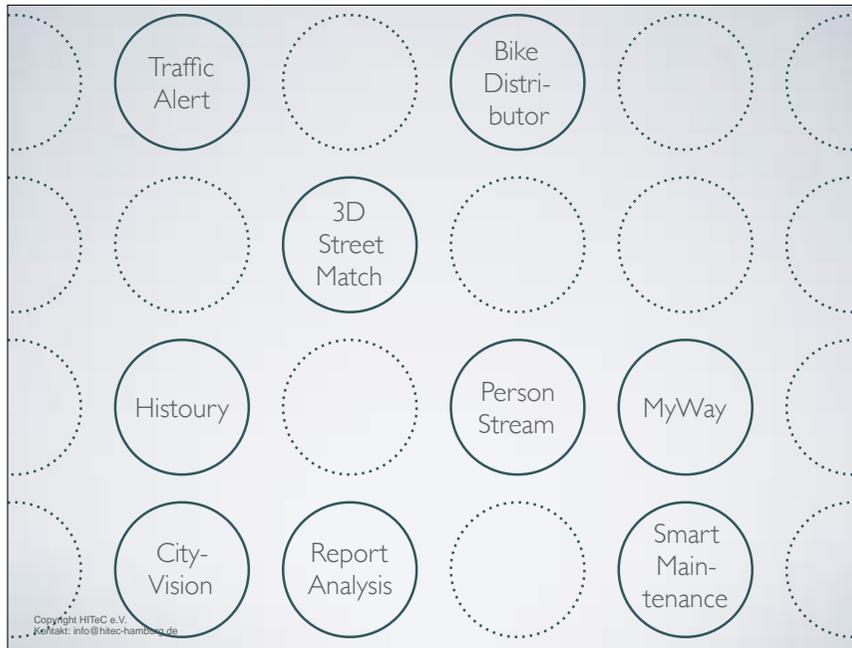
- Umfasst zahlreiche Teildisziplinen der KI
 - Objekterkennung, Szenenanalyse, Planen, Lernen
- Kompliziertes Zusammenspiel von Sensorik und Verarbeitungsverfahren
 - z.B. automatisches Erkennen von Schildern
- Autonomes Fahren kommt
 - Leistungsfähigkeit kann mit Erfahrungen wachsen

WICHTIGE EIGENSCHAFTEN VON KI-METHODEN

- Integration heterogener Daten in kohärentem semantischen Rahmen
- Lernen anhand von (vielen) Beispielen und Statistiken
- Entdecken von Strukturen
- Defizite bei Transparenz und Erklärungsverfahren

3

IDEENSKIZZEN & VORSTUDIEN



VORGEHEN

1. **Bildverarbeitung**/Monitoring von Objekten
2. Beschreibung von möglichen **Ereignissen und Objekten** im Straßenverkehr; verknüpft **mit zeitlichen Beschränkungen** (Regeln)
 - Mithilfe einer **Ontologie**, SWRL
3. **Scenior** (scene interpretation by ontology-based rules) **übersetzt** die Wissensbasis mit den zeitlichen Bedingungen in **ein Regelsystem**, welches die Realdaten aufgrund der modellierten Ereignisse auswertet

LEISTUNGEN DER BILDVERARBEITUNG

Verkehrsteilnehmer/Fahrzeuge werden

- **identifiziert** (PKW, Krankenwagen, Transporter) z.B. durch zuvor trainierte neuronale Netze und
- **getrackt** durch Monitoring-Systeme.



IN DER WISSENSBASIS...

erfolgt die Beschreibung von Primitiv-Ereignissen (Vorwärts-, Rückwärtsfahren, Stillstand) der Verkehrsteilnehmer, wie z.B.

```
Drive
hasDirection exactly 1 Direction
hasRoadUser exactly 1 RoadUser
hasLane exactly 1 Lane
hasZone exactly 1 Zone
hasVelocity exactly 1 Velocity
```

zudem durch Startzeit, Endzeit.

SO SIEHT EIN PRIMITIV-EREIGNIS IN DEN KAMERADATEN AUS...



Drive (car1, lane2, zone1)

Start: 00:12:34

End: 00:12:36

IN DER WISSENSBASIS...

...erfolgt die Beschreibung von Objekten, wie z.B. der Straßenführung:

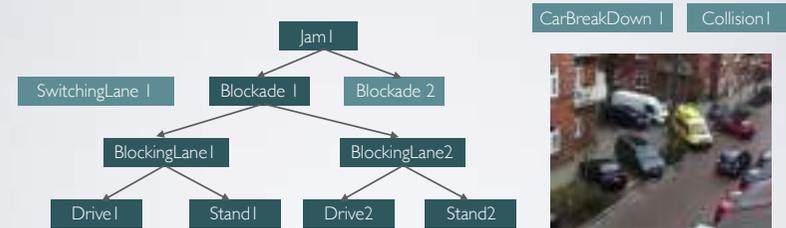


Kreuzung mit Aufteilung in Spuren, die in Zonen unterteilt sind.

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

IN DER WISSENSBASIS...

- ...werden bekannte Verkehrssituationen modelliert, die aus den Primitiv-Ereignissen bestehen.



Drive-Bedingung: zeitlich direkt vor Stand, auf gleicher, Spur und mit gleichem Fahrzeug

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

SCENIOR



Es handelt sich um eine gegenseitige Blockade.

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

TRAFFIC ALERT

- ➔ Echtzeitauswertung der Bilddaten der Verkehrskameras durch Bildverarbeitung, Monitoring und Scenior
- ➔ Mit dem Ergebnis der Szeneninterpretation können Verkehrsplaner informiert werden, damit sie Optimierungen im Verkehrsfluss einleiten

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



PROJEKTIDEE

⚡ HERAUSFORDERUNGEN

- Steigende städtischen Umweltbelastungen fordern nachhaltige, attraktive Alternativen zur Fortbewegung mit dem PKW.
 - Eine Alternative: Bike-Sharing
- Dazu müssen Bikes zuverlässig allerorts zur Verfügung stehen

🎯 ZIEL

Bedarfvorhersage zur Verbesserung der Disposition

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

VORGEHEN

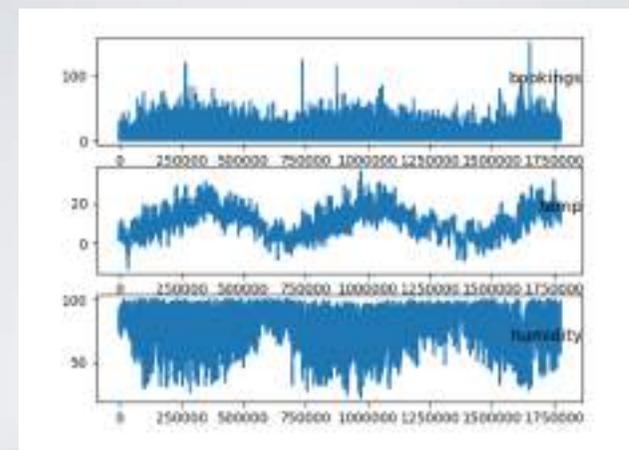
Buchungen der Stadträder an best. Stationen werden korrelieren mit

1. Wochentag/Feiertag
2. Wetterinformationen
 - Temperatur
 - Niederschlag

➔ Vorhersage des Radbedarfs mit Hilfe eines Neuronales Netzes

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

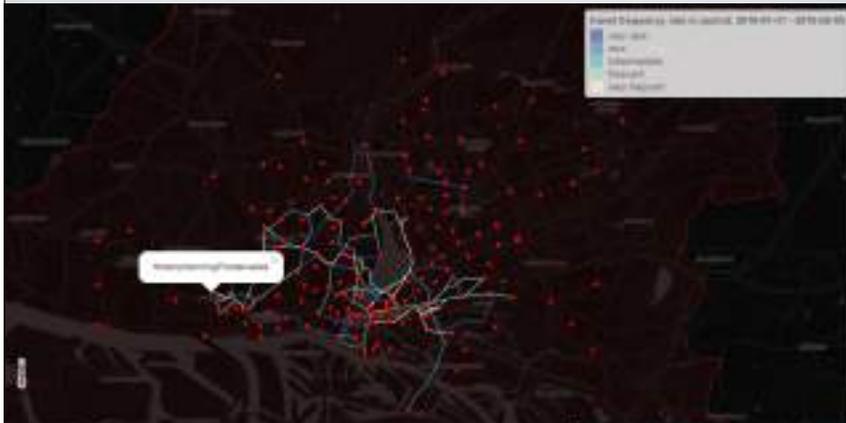
DATEN



Umfang 2 Jahre

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

HÄUFIGSTE STADTRADTOUREN VISUALISIERT ...



Basierend auf den Open Data Buchungsdaten der Deutschen Bahn

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

STADTRADROUTEN MIT WETTERDATEN ANGEREICHERT ...



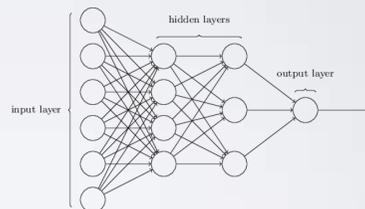
mit Open Data des DWD

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

NEURONALES NETZ

Input für das Neuronale Netz

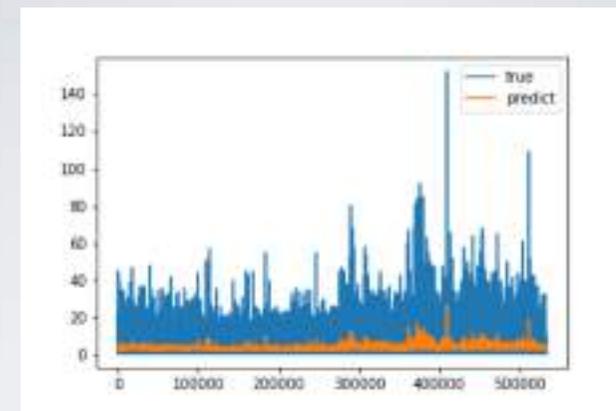
- Stadtradstation, Temperatur, Luftfeuchtigkeit, Wochentag, Feiertag und Anzahl der Buchungen
- Netz-Konfiguration:
 - 3-Layer Neuronales Netzwerk
 - Sequentiell angeordnete Layer
 - Fully Connected Layers
 - Aktivierungsfunktion: ReLu
 - 10 Epochen a 1000 Input-Daten



70% Trainings- und 30% Testdaten

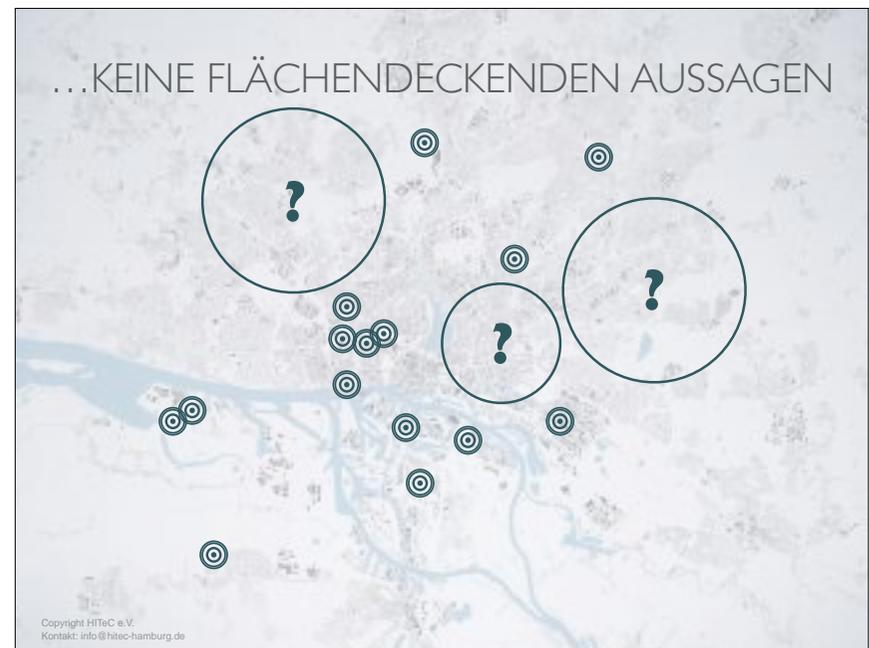
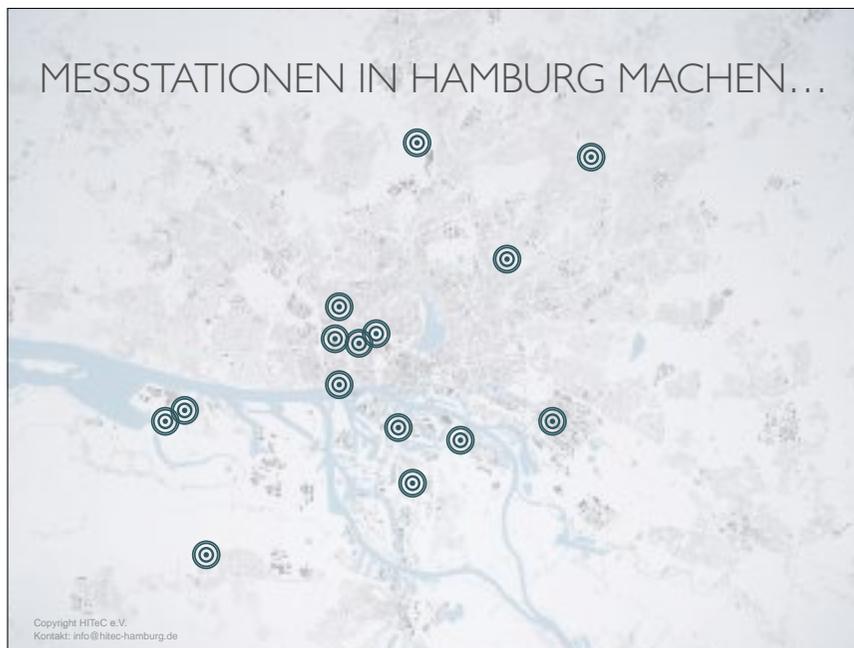
Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

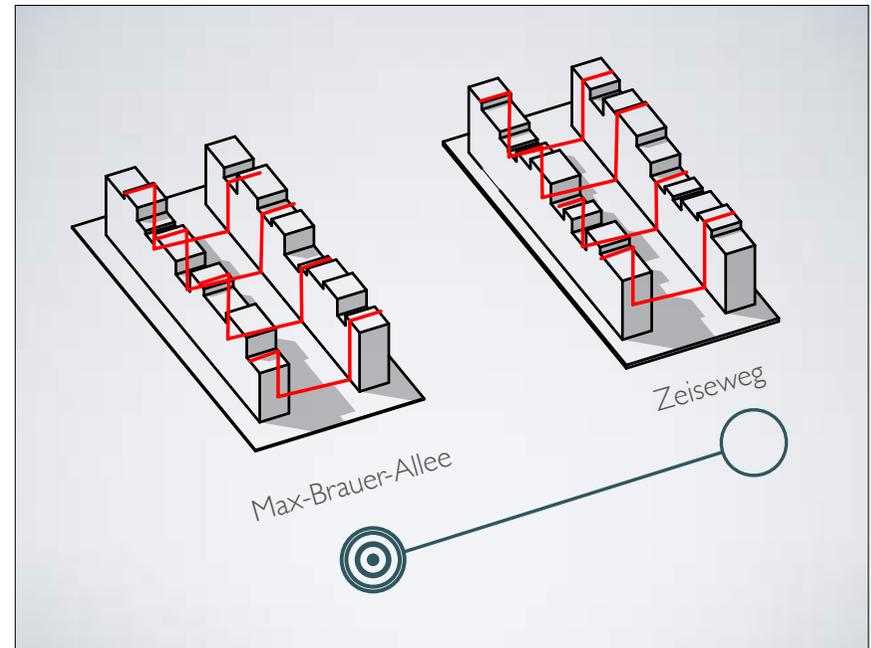
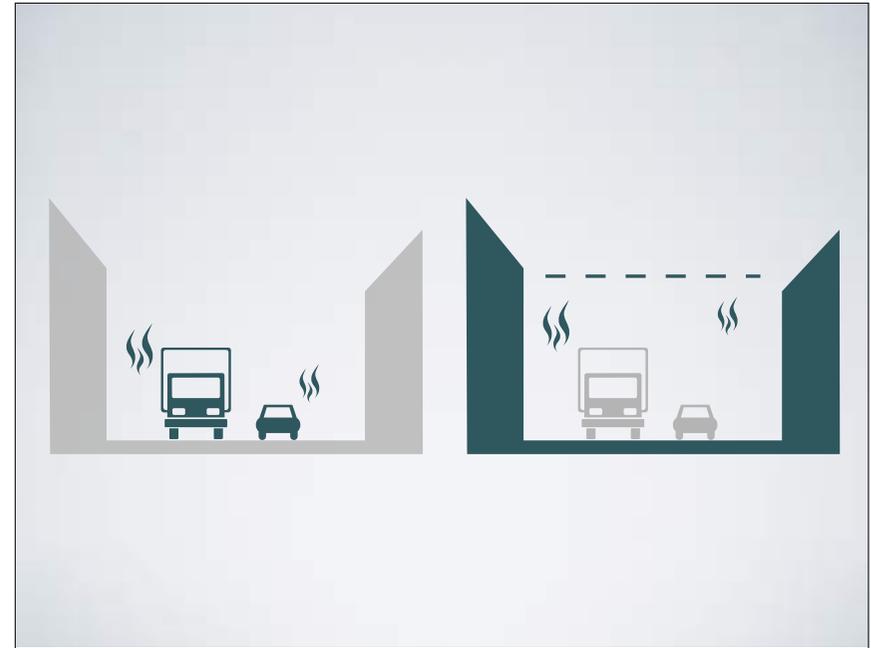
VORHERSAGE



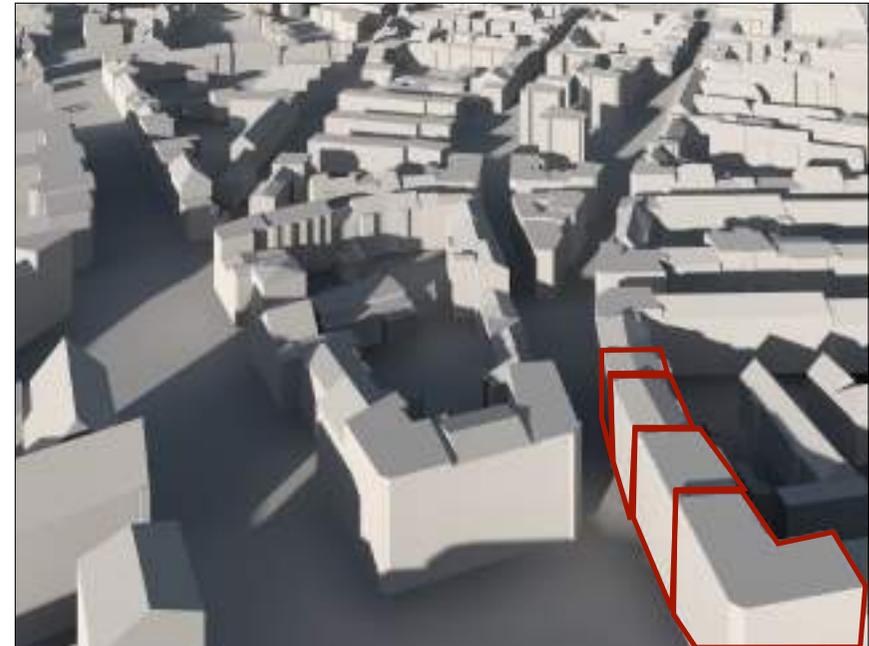
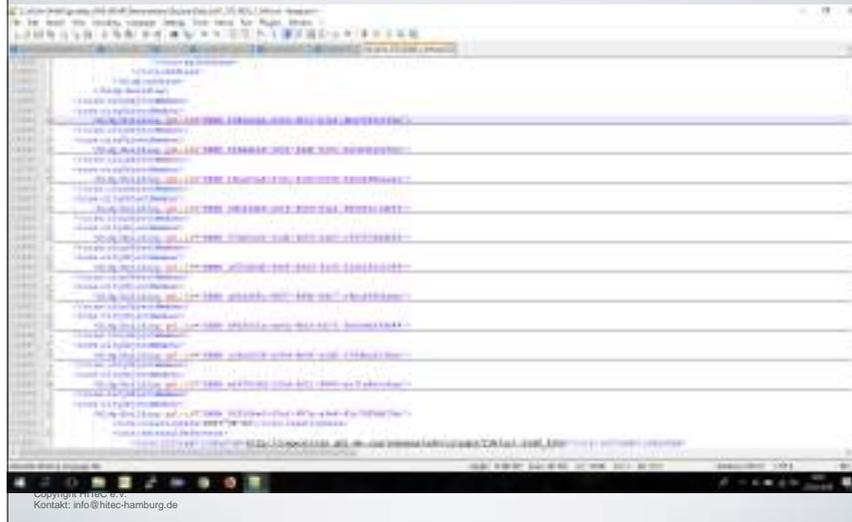
Zu wenig Daten um gute Vorhersagen zu machen. Durch Hinzufügen eines weiteren Features, ggf. Eventdaten bessere Ergebnisse.

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

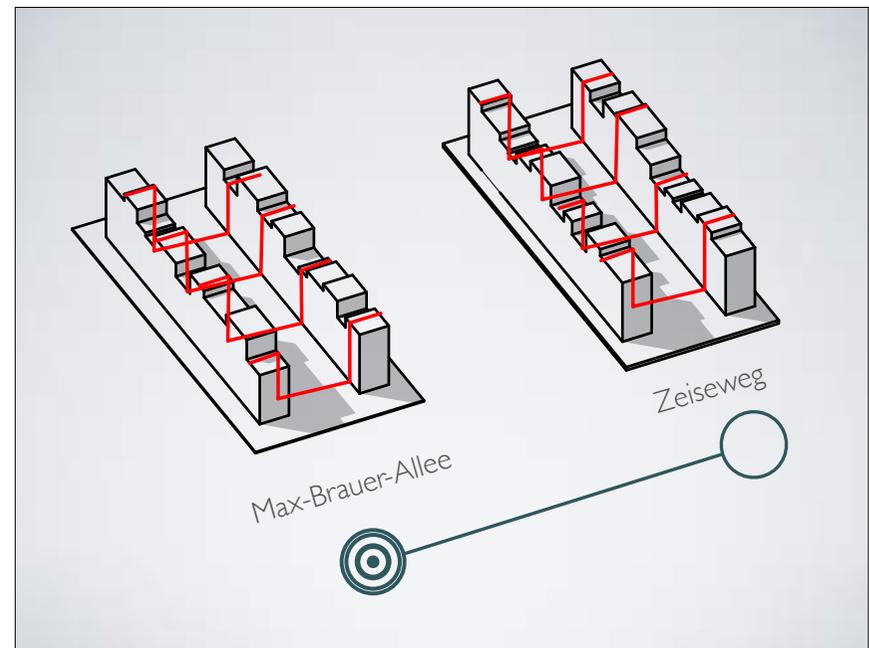
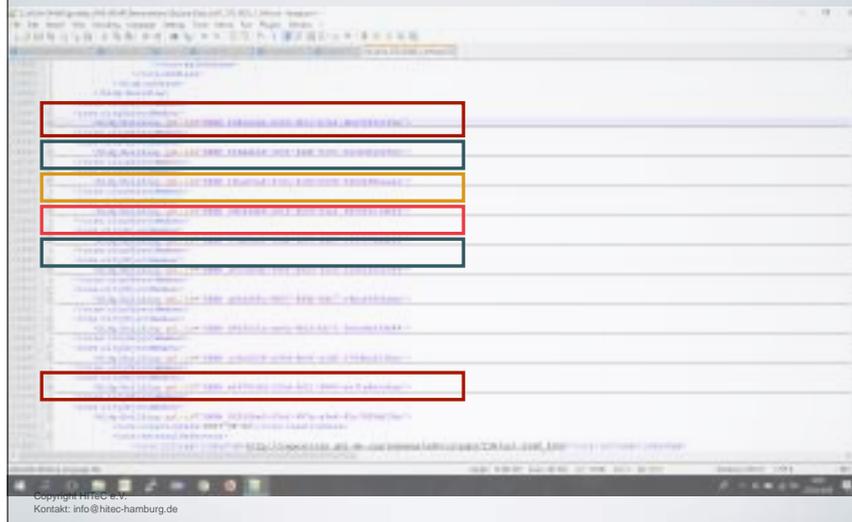




DATENSATZ DES 3D MODELLS



DATENSATZ DES 3D MODELLS



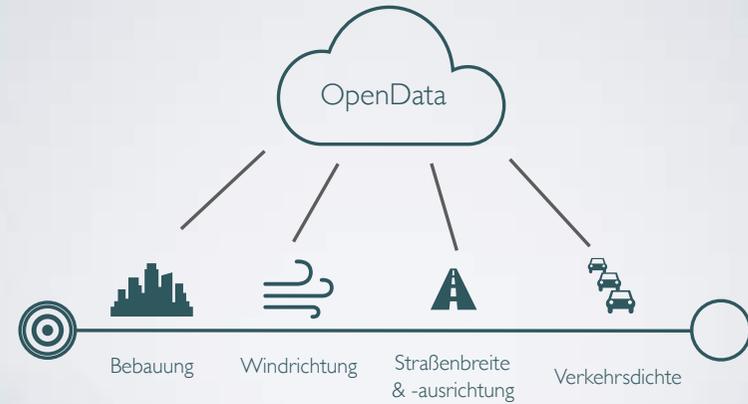
BEISPIEL EINES MERKMALSVEKTORS:

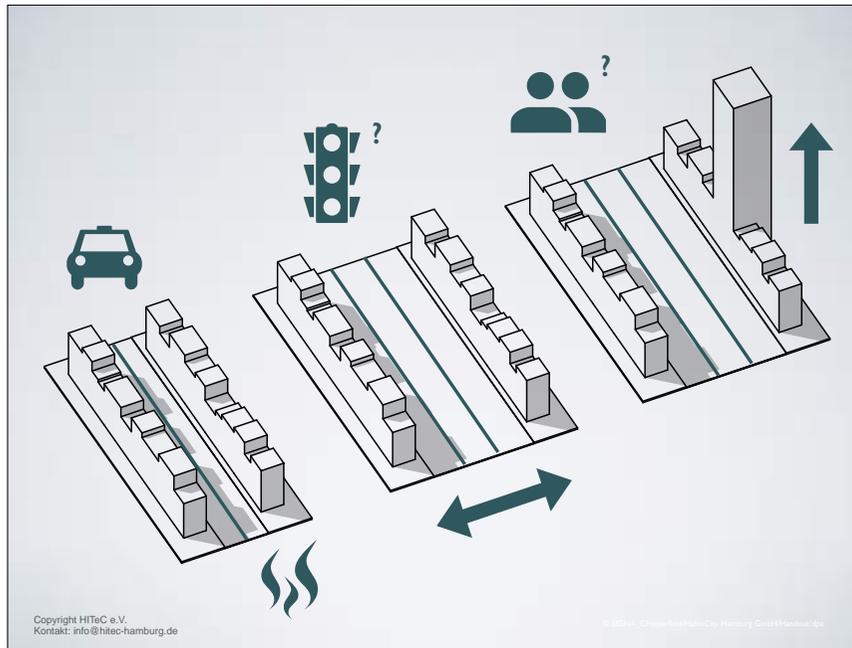
Long	Lat	SR1	SR2	SR3	SR4	SR5	SR6	SR7	SR8
SR9	SR10	SL1	SL2	SL3	SL4	SL5	SL6	SL7	SL8
SL9	SL10	StreetWidth	CK-Length	Orientation	Traffic				
9.943290587610491		53.54637680428949		19.084	17.163	14.013	0.0		
0.0	0.0	0.0	0.0	0.0	5.104	17.163	0.0	0.0	0.0
19.99754912681782		10.30391716085444		91.54909983203319					
20000									

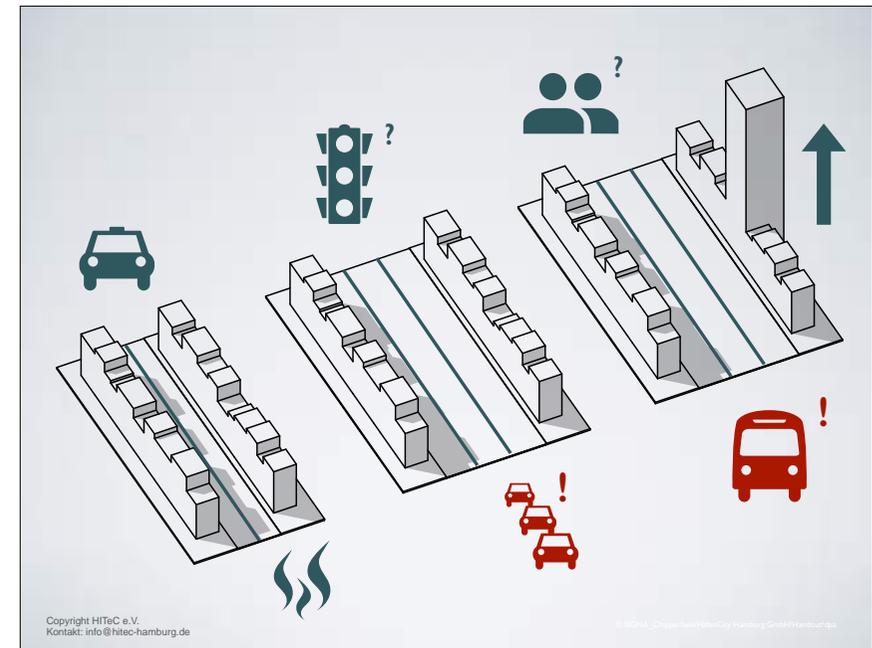
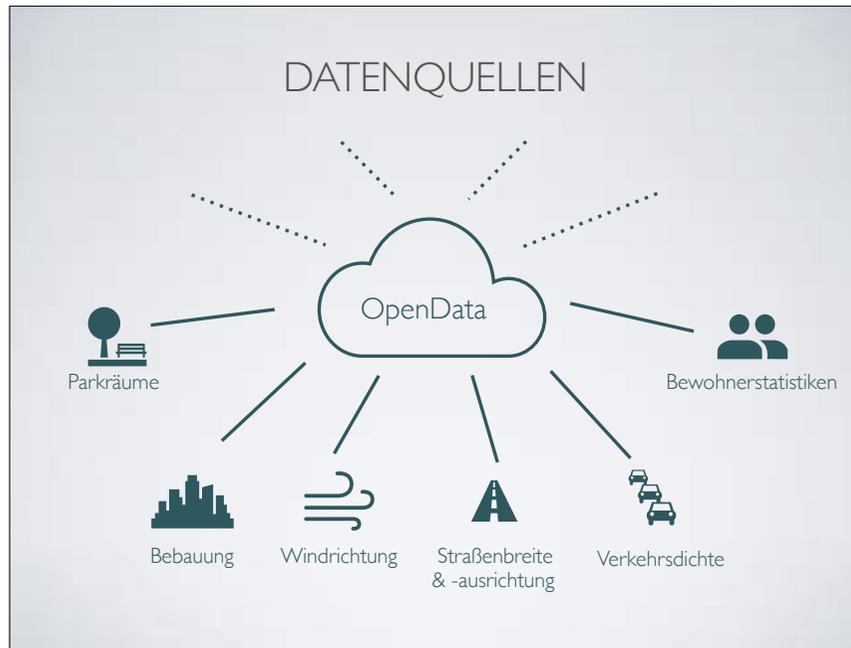
SL, SR = Höhen der beiden Seiten eines Straßenzuges

Copyright HITEC e.V.
Kontakt: info@hitec-hamburg.de

MEHRDIMENSIONALER MUSTERVERGLEICH







HISTOURY

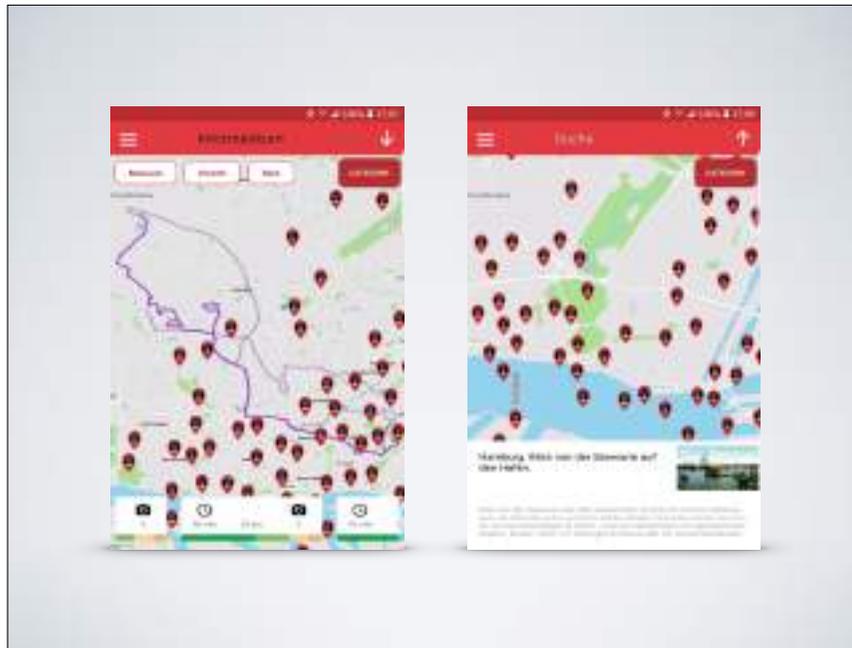
👁️ BEOBACHTUNGEN

- **Viele Touristen** in der Hansestadt
- jährlich **neue Studierende**, die die Stadt kennenlernen wollen
- **historische Fotos** zu Orten und Sehenswürdigkeiten stehen für Radbummler **kaum zur Verfügung**

💡 IDEE

Mit einer **APP** wird dem Ortsfremden Hilfe bei der **Routenplanung** gegeben, bei der **Radtour** wird der Fahrer auf interessante historische **Sehenswürdigkeiten hingewiesen**

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



FEATURES VON HISTOURY

- Erkunde vielfältige historische Sehenswürdigkeiten in deiner Nähe 
- Planung der Fahrradrouten 
- Gib persönliche Interessen ein 
- Werde über in der Nähe befindliche Sehenswürdigkeiten informiert 

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de

UMSETZUNG

- Wurde als demonstrierbarer **Prototyp am Fachbereich Informatik** implementiert
- Die Umsetzung erfolgte von einer **7-köpfigen Gruppe von Studierenden** im Projektzeitraum
- Das Ergebnis wurde im Januar 2018 erfolgreich demonstriert.
- Datenquellen:
 - **efoto-hamburg.de**
 - Karten
 - Fahrradtouren

Copyright HITeC e.V.
Kontakt: info@hitec-hamburg.de



PERSONSTREAM

Steigerung der Attraktivität und bedarfsgerechte Ausrichtung von ÖPNV/ Reiserouten

Motivation für private PKW-Fahrten reduzieren

Dazu sind Informationen über die Ziele und Bedürfnisse von Fahrgästen nötig.

Im Moment: Punktuelle, analoge Zählung und Befragung der Kunden

- sehr aufwendig, gibt nur wenig Einblicke in Kundenbedürfnisse
- Bewegungsdaten sollen von Personen erfasst, analysiert und anschaulich präsentiert werden

ERFASSEN VON BEWEGUNGSDATEN

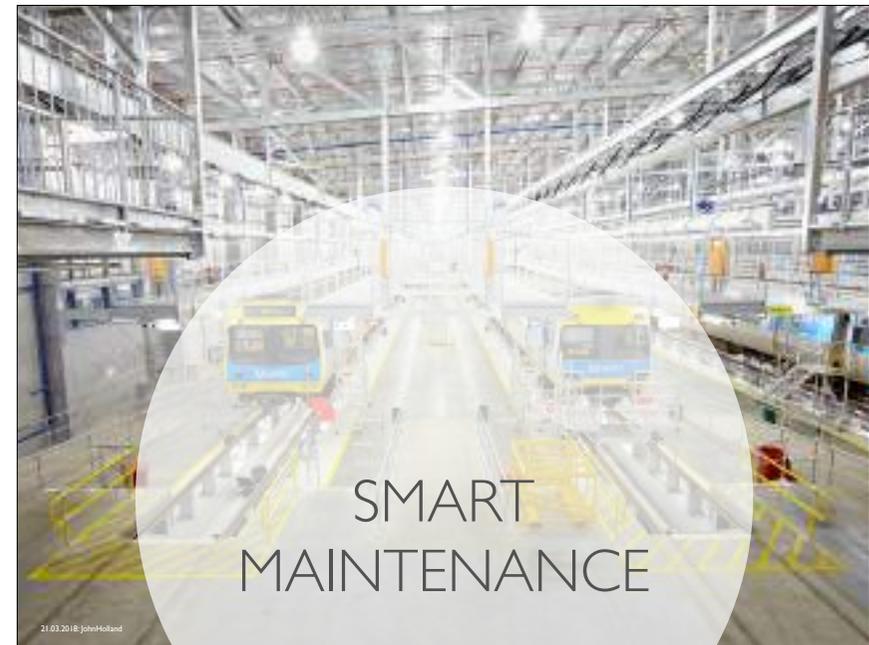
Erfassen von Bewegungsdaten ist heute möglich durch Auswertung der flächendeckenden Nutzung von mobilen Endgeräten:

- Mobilfunkdaten von Kunden (anonymisiert und aggregiert)
- Lokale Hotspots/WLAN eingeloggte Geräte (anonymisiert und aggregiert)
- Ausnutzung von Daten durch APP-Nutzungen

Aber auch durch die quantitativen Auswertung von Passagierströmen durch Kameradaten

MUSTERERKENNUNG IN BEWEGUNGSDATEN

- Mithilfe einer Clustering-Komponente (K-Means oder ein NN) können häufige Muster in den Bewegungsdaten gefunden werden
 - Durch die zusätzliche Korrelation mit Kontextdaten kann das Muster dann als z.B. Berufsverkehr, Einkaufstour, Veranstaltungsbesuch gedeutet werden.
- ➔ So können Shop-Angebote oder Werbetafeln je nach erwarteten Mustern angepasst werden, z.B.
- Was braucht ein Pendler oder geschäftlicher Fluggast auf dem Weg zu Arbeit?
 - Was könnte ihn auf dem Rückweg ansprechen?



SMART MAINTENANCE (1)

- Wartung von Verkehrsmitteln erfolgt in der Regel
 - in festen Wartungsintervallen
 - bei Störungen
- Durch vorhersehende Wartung (Preventive Maintenance)
 - können Störungen vermieden werden
 - Wartungsarbeiten an den Bedarf angepasst werden

SMART MAINTENANCE (2)

VISIONÄRER LÖSUNGSANSATZ:

- Internet-der-Dinge (Internet of Things, IoT)
 - Komponenten besitzen Sensoren
 - Komponenten melden laufend ihren Zustand
- Lernen von störungsträchtigen Situationen (Systemzuständen)
 - Speichern von Big Data
 - Überwachtes Lernen mit KI-Methoden
- Laufendes Analysieren von aktuellen Sensordaten
 - Mustervergleich mit gelernten Störungssituationen
 - Anomalieerkennung

REPORT ANALYSIS

- Auswerten von gesammelten textuellen Berichten über
 - Unfälle
 - Schadensfälle
 - Anzeigen
- Ermitteln von gemeinsamen Ursachen und Begleitumständen
 - bei schlechter Sicht
 - an neuralgischen Verkehrsknoten
 - an Feiertagen

REPORT ANALYSIS (2)



TEXTUELLE UNFALLBERICHTE LEIDER NOCH NICHT OPEN DATA

LÖSUNGSANSATZ

- Sprachliche Analyse der Texte
 - semantische Einordnung mit KI-Sprachanalyse
- Entdecken von Zusammenhängen
 - Clusteranalyse mit NN
 - Mustererkennung mit gelernten Klassifikatoren

Copyright HITeC e.V.
Kontakt: info@hittec-hamburg.de

4

ERFAHRUNGEN & AUSBLICK

PROJEKTE UND KI-VERFAHREN

TrafficAlert:	Ontologie und Objekterkennung
BikeDistribution:	Statistik, Neuronale Netze
3D-StreetMatch:	Mustererkennung
CityVision:	Ontologie und Mustererkennung
PersonStreams:	Statistik, Neurone Netze, Mustererkennung
History:	Handlungsplanung und Mustererkennung
MyWay:	Handlungsplanung
ReportAnalysis:	Textverarbeitung
SmartMaintenance:	Big Data, NN, Mustererkennung, Ontologie

Copyright HITeC e.V.
Kontakt: info@hittec-hamburg.de

NUTZEN VON KI-VERFAHREN (I)

- Eine Vielzahl von Datenquellen stehen zur Verfügung
 - Leider noch nicht alle Open-Data (z.B. Personenbewegungen)
- KI-Methoden sind realisiert und als Werkzeuge verwendbar
 - R, TensorFlow, SKLearn, Weka, Protégé, JESS, ...
- KI unterstützt Integration von vorhandenen Datenquellen
 - ÖPNV, Schadstoffquellen, Ereignisse, Wetter, Wochentage, Jahreszeit
- KI bietet Mustererkennung und Klassifikation
 - Erkennung von Verkehrsteilnehmer
- KI ermöglicht Vorhersagen
 - Schadstoffbelastung bei Bauplanung

Copyright HITeC e.V.
Kontakt: info@hittec-hamburg.de

NUTZEN VON KI-VERFAHREN (2)

- Prädiktive Wartung
 - Verkehrsmittel
 - Komplexe Szeneninterpretation
 - Vorausschau beim autonomen Fahren
 - Agententechnologie
 - Internet of Things für Verkehrsmittel und städtische Bauten
 - Visualisierung
 - Erkennen von Zusammenhängen in großen Datenmengen
- ➔ Nutzen für die Wirtschaft, Stadtplanung, Bürger

ANMERKUNGEN ZU DEEP LEARNING

- Verwendet Künstliche Neuronale Netze (NN)
- Eines von mehreren KI-Lernverfahren
- Hauptanwendung bei Mustererkennung, Klassifikation, Bildverarbeitung
- Supervised: Positive und negative Beispiele notwendig
- Unsupervised: Erkennung von Zusammenhängen
- Befähigt durch: Rechenleistung, Datenmengen, Cloud

HERAUSFORDERUNG FÜR ANWENDUNGSENTWICKLUNG

- Datenhungrig: große Mengen von Trainingsdaten
 - Regelmäßiges Datensammeln im Bereich Open-Data!
- Domänenspezifisch: Transfer ist problematisch
 - Lernen muss für einzelne Bereiche durchgeführt werden
- Undurchsichtig: die Vielzahl der Parameter eines NN können nicht interpretiert werden, keine Erklärbarkeit
 - Aber notwendig zur Begründung von Entscheidungen
- Oberflächlich: kein semantisch strukturiertes Wissen

Für umfassende Aufgaben (z.B. Planungssysteme, vorausschauendes Fahren) sind hybride Ansätze erforderlich (Mustererkennung plus Reasoning)

DURCHFÜHRUNG VON KI-PROJEKTEN

- Verfahren der KI liegen als Werkzeuge vor
- KI-Projekt besteht aus:
 - Akquisition der Daten
 - Datenaufbereitung der Daten
 - Auswahl und Kombination von KI-Verfahren
 - Modellierung des Problems
 - Experimentierphase
 - Architektur und Integration in das vorhandene IT-System

ROLLEN IN EINEM PROJEKT

- AnwenderInnen
 - Beispiele: BürgerInnen, PlanerInnen Hafen, Flughafen
- Domänenexperte
 - Verkehr, Geodaten, Stadtplanung
- KI-ExpertInnen
 - NN, Ontologien, Tools
- Software-Haus
- Integration, Benutzerinteraktion

SCHLUSSWORT

- KI-Verfahren angewandt auf Open-Data führen zu effektiven Anwendungen
- Projektbeschreibungen von ODAKI liefern Startpunkte
- Kombinierte KI-Verfahren erlauben komplexe Planungs- und Monitoringaufgaben
- HITeC plant die Durchführung des Projekts Traffic Alert

5

DISKUSSION