



## Forschungsprojekt

# Analyse von mikroskopischen Nutzungsdaten als Grundlage für die Entwicklung flexibler Tarife

Laufzeit: Januar 2016 bis Dezember 2016

Zuwendungsgeber:  
Land Hessen

### HINTERGRUND UND ZIEL

EFM-Systeme erzeugen im Rahmen der Abrechnung Nutzungsdaten, die sämtliche in Anspruch genommenen Leistungen der Kunden dokumentieren. Bei diesen Nutzungsdaten handelt es sich um Beobachtungsdaten, die das Verhalten der gesamten Nutzergruppe über längere Zeiträume erfassen. Trotz der Tatsache, dass nur derjenige Ausschnitt des Mobilitätsverhaltens abgebildet wird, der in Zusammenhang mit dem EFM-System steht, können diese Daten für verschiedene Planungsaufgaben als Datengrundlage wertvoll sein. EFM-Systeme ermöglichen darüber hinaus die Weiterentwicklung bzw. Ergänzung des bestehenden Tarifportfolios im ÖPNV durch neue Tarifprodukte, die sich durch eine Flexibilisierung der nutzungsabhängigen Fahrpreisbildung auszeichnen. Die Zielsetzung des Projekts bestand darin, die Nutzungsdaten eines EFM-Systems systematisch auszuwerten und insbesondere Nutzungsmuster zu erkennen, anhand derer sich Nutzerprofile ableiten lassen, die eine Entwicklung bedarfsgerechter, flexibler Tarifprodukte für EFM-Systeme im ÖPNV ermöglichen sollten.

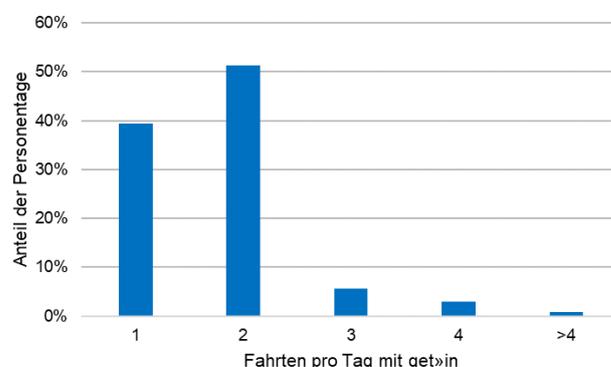
### DATENGRUNDLAGE

Im Rahmen dieses Projekts wurden die Nutzungsdaten des get»in-Systems analysiert, das von 2001 bis 2013 in den benachbarten Städten Hanau und Bruchköbel als chipkartenbasiertes Check-In-Check-Out-System der dritten Stufe nach VDV-Kernapplikation (automatisierte Fahrpreisfindung) betrieben wurde. Eine tarifliche Besonderheit des Systems war dabei das Best-Preis-Prinzip. Aufgrund dessen wurde dem Kunden im Gültigkeitszeitraum einer Zeitkarte maximal der Preis dieser Zeitkarte in Rechnung gestellt. Sobald also beispielsweise die Umsätze eines Kunden innerhalb eines Monats durch einzelne Fahrten den Preis einer Monatskarte überschritten, konnte er das System bis zum Monatsende ohne weitere Kosten nutzen. Daher war das get»in-System nicht nur für Gelegenheitsnutzer, sondern auch für Stammkunden und häufige Nutzer geeignet.

### METHODISCHES VORGEHEN UND ERGEBNISSE

Da der primäre Zweck der Transaktionsdaten des Systems die Abrechnung der in Anspruch genommenen Leistung war und die originäre Datenstruktur sich nicht direkt für statistische Auswertungen eignete, mussten die Daten zunächst einer umfangreichen Aufbereitung und Plausibilisierung unterzogen werden. Im Anschluss daran konnten zunächst deskriptive Analysen der Daten vorgenommen werden. Dadurch konnten bereits verschiedene Erkenntnisse generiert werden.

Beispielhaft ist in der folgenden Abbildung zu erkennen, dass an etwa 40% der Nutzungstage nur eine Fahrt pro Nutzer vom get»in-System aufgezeichnet wurde. Erklärungsansätze für diese Beobachtung sind durchaus vorhanden, aber der große Anteil überrascht dennoch. Entsprechende Anteile ähnlicher Größenordnung konnten mittlerweile auch in anderen Datenquellen identifiziert werden.



Als vielversprechender Ansatz für die Identifizierung der Nutzerprofile wurde das mathematische Verfahren der Clusteranalyse ausgewählt. Es hat insbesondere den Vorteil, dass Gruppen mit ähnlichen Merkmalsausprägungen ohne Vorkenntnisse gebildet werden können und daher keine willkürliche Gruppenbildung und anschließende Zuordnung durch den Datenanwender erfolgt. Im Rahmen des Projekts wurden die Daten zunächst auf vorhandene zeitliche Muster im Tagesablauf untersucht. Die Ähnlichkeit der einzelnen „Nutzertage“ wurde in diesem Zusammenhang im Rahmen einer Sequenzdatenanalyse bestimmt, deren wesentliche Arbeitsschritte im Folgenden skizziert sind:

#### 1. Nutzungsdaten in binäre Sequenzen überführen:

Zu diesem Zweck wird der Tag in Intervalle gleicher Länge aufgeteilt und eine Sequenz gebildet, in der eine „1“ gleichbedeutend mit einer Nutzung in diesem Zeitintervall und eine „0“ gleichbedeutend mit „keine Nutzung in diesem Zeitintervall“ ist.

#### 2. Abstandsmaß für die Sequenzen berechnen:

Im konkreten Anwendungsfall wurde die Levenshtein-Distanz als Distanzmaß gewählt, welche die minimal notwendige Anzahl der bitweisen Operationen auf eine Sequenz A ermitteln, die notwendig ist, um diese in eine Sequenz B zu überführen.

#### 3. Clusterverfahren anwenden:

Hier wurden das Single-Linkage-Verfahren zur Eliminierung von Ausreißern und das Ward-Verfahren für die eigentliche Clusterbildung gewählt, die der Gruppe der hierarchischen Clusterverfahren zugeordnet werden.

Im Rahmen der kurzen Projektlaufzeit konnte das geplante Vorgehen realisiert werden. Leider konnten in den Nutzungsdaten der Gelegenheitskunden keine entsprechenden aussagekräftigen Muster entdeckt werden.

Das Projekt wurde vom Land Hessen im Rahmen der HOLM-Förderrichtlinie „Innovative Projekte in Logistik & Mobilität“ gefördert und im Verbund mit der Rhein-Main-Verkehrsverbund Servicegesellschaft mbH umgesetzt.