# Problemstellung

Destatis, Umweltbundesamt, Arbeitsagentur, BMVI, Kraftfahrtbundesamt

* SGV macht knapp 80% des Gesamtdeutschen Güterverkehrs aus (neben Bahn, Schiff und Luftfracht)
* 3643 Mrd. Tonnen beförderte Güter bei 479 Mrd. Tonnenkilometern Beförderungsleistung 2017
* SGV wuchs zwischen 1991 und 2016 um 89 %!
* Umweltbelastung entsprechend hoch (Komponenten: Fahrleistung = # gefahrener km, Geschwindigkeit und Fahrzeugart.
* Moderne Fahrzeuge haben heute zwar einen geringeren Schadstoffausstoß, trotzdem nehmen die Emissionen nicht im gewünschten Maße ab (Klimaziele)
* Derzeit massiver Kraftfahrermangel
* Ca. 151 Mio. Leerfahrten und 6.504,8 Millionen Kilometer 2016
* Kurzfristprognose des Sendungsaufkommens derzeit nur mit internen historischen Daten
* Planung vor wichtigen Feiertagen (Ostern, Pfingsten, Weihnachten etc.) 🡪 Prognosehorizont ca. drei Wochen
* Einsatzplanung von Personal in der kurzen Frist 🡪 Prognosehorizont ca. 1-7 Tage

# Projektziel

* Hinzuziehen externer Größen (Wetter, Feiertage, Ferienkalender, Branchenkonjunkturdaten/ BIP, Unfall, Streiks/Ausfälle in den anderen Verkehrsmitteln?)
* Prognosemodelle anhand statistischer Methoden ML/KI
  + ANN !!
  + (S)ARIMA
  + Random Forest
  + SVR
* Prädiktiver Mehrwert durch ML?

# NAGEL-Group

* Spezialist für Lebensmitteltransporte
* Standort Nürnberg fokussiert auf Warehousing und Stückgut national
* europäische Plattform für Tschechien, Österreich, Großbritannien und Italien
* LTL- und FTL

## Datensatz

* ABS\_REL: Absende-Terminal-Nr. (90 = Nürnberg, etc.)
* Ziel\_REL: Ziel-Terminal/-Relation
* akdid2: Absenderkunden-ID (PLZ-V…)
* ekdid2: Empfangskunden-ID
* vk\_strom2: SG = Stückgut, Dir = Direktsendung an Verbraucher, DirVNL = ?
* SEN\_ART: Sendungsart (A=Ausgang, N=Nahverkehr)
* LST\_D: Datum
* Kw = Kalenderwoche
* Wotag = Wochentag
* Quartal
* Feiertagswoche (binär)
* Event: Feiertagsmerkmal (Pfingsten-3 = 3 Wochen vor Pfingsten)
* gewicht: Gewicht in kg
* Wetterstations-ID + Temp: Gemessene Temperatur an der Zielwetterstation in Celsius
* Alle Dateneinträge haben die gleiche Dimension, d.h. 30 Variablen mit je 472.137 Beobachtungen
* Es liegen 49 fehlende Werte in der Spalte “gewicht” vor.
* Es gibt 5.525 verschiedene Absender.
* Es gibt 15.576 verschiedene Kunden/Empfänger.
* Es liegen 13 Events vor.
* Klarer Unterschied Werktag – Wochenenden, Feiertage
* Die Auslieferung erfolgt i.d.R. T+1 und bei Feiertag entsprechend länger
* Ein LKW fasst ca. 12 Tonnen

# Gliederung

1. Einleitung
2. Theorieteil
   1. State of the art: Freight Demand Forecasting (for Truckload Shipping)
      1. Chow et al. 2010:
         * Einteilung in (u.a.) vehicle-based und commodity-based models
      2. Zhou, Heimann & Clausen (2006):
         * ARIMA und multilayer feedforward NN
         * NN trend- und saisonbereinigt am vorhersagestärksten
      3. Moscoso-Lopez et al. (2016):
         * Freight forecasting of fresh vegetables in Port of Algeciras (Schifffahrt!)
         * Forecast of daily weight of the freight with one week in advance
         * Multilayer feedforward NN und SVR
         * SVR results in better MAPE
      4. Generell: Exp. Glättung, ARIMA (x), **Manuelle Planung;** Aufgabe:Externe Daten
   2. Data Understanding

Deskriptive Statistiken, Visualisierung, Niveau, fehlende Werte, „Auflösung“, Ranges, Ausreisser….

* 1. Data Preparation

Ausreisser bereinigen, Granularitäten anpassen, Skalieren

* 1. Modeling

Feature Selection:

* Aufgabe: Externe (öffentlich zugängliche Daten) mit einbeziehen, da wichtige wirtschaftliche, demographische und geographische Kenngrößen einen Einfluss haben könnten
  + Nested models (nur interne vs. Interne + externe Daten)

Aufteilung Training <-> Testset

Auswahl der Methode

* 1. Evaluierung

Testmetriken (AUC, Konfusionsmatrix, (R)MSE, MA(P)E…)

1. Empirischer Teil *Umsetzung des in der Theorie Beschriebenen*
   1. Business Understanding

Beschreibung Prozess, Nagel…

* 1. Data Understanding
  2. Data Preparation
  3. Modeling
  4. Ergebnisdarstellung

1. Ergebnisdiskussion
2. Zusammenfassung & Ausblick
   1. Schwächen der Arbeit, Weitere Forschungsperspektiven