

Analysemethodik

Um aussagekräftige und sinnvolle Ergebnisse bei der Datenanalyse zu erzielen, wird der **Data Science Lifecycle** als Struktur genutzt. Dieser Lifecycle besteht aus mehreren Phasen: Business Understanding, Data Mining, Data Cleaning, Data Exploration, Feature Engineering, Predictive Modeling und Data Visualization. Jede Phase ist entscheidend, um die gesammelten Fahrdaten systematisch zu aggregieren und zu analysieren.

1. **Business Understanding:** Der erste Schritt im Data Science Lifecycle ist das Verständnis des Geschäftskontexts und der Ziele der Analyse. In dieser Untersuchung ist das Ziel, die Effizienz und Nachhaltigkeit von Elektrofahrzeugen (EVs) im Vergleich zu Fahrzeugen mit Verbrennungsmotoren (ICEs) zu bewerten. Dies wird durch die Analyse von Fahrdaten realisiert, die in Hamburg aufgezeichnet wurden.

Die zentralen Fragestellungen betreffen die Energieeffizienz unter realen Fahrbedingungen und den Vergleich mit den Normwerten des Worldwide Harmonized Light Vehicles Test Procedure (WLTP).

2. **Data Mining:** In der Data Mining-Phase werden die Fahrdaten gesammelt und aufbereitet. Für diese Untersuchung wurden Fahrdaten sowohl von einem Elektrofahrzeug (Mokka E) als auch von einem Fahrzeug mit Verbrennungsmotor (Golf 8) in Hamburg erhoben. Die Datenerhebung erfolgte werktags über zwei Tage, wobei jeweils 30 Fahrten pro Fahrzeug aufgezeichnet wurden. Diese Fahrdaten umfassen unter anderem Zeitstempel, geografische Koordinaten, Geschwindigkeit, zurückgelegte Distanz, Höhe, Beschleunigung sowie Kraftstoffverbrauch und CO₂-Emissionen.

3. **Data Cleaning:** Die Datenbereinigung ist ein kritischer Schritt, um die Qualität der Daten sicherzustellen. Hierbei werden Fehler und Inkonsistenzen in den Rohdaten beseitigt.

Dies umfasst:

- Entfernung von Ausreißern: Identifikation und Entfernung von Datenpunkten, die signifikant von der Norm abweichen.
- Behandlung fehlender Werte: Interpolation oder Entfernung von Datensätzen mit fehlenden Werten.
- Korrektur fehlerhafter Daten: Überprüfung und Korrektur von Datenpunkten, die durch Messfehler oder andere Anomalien entstanden sind.

4. **Data Exploration:** Nach der Datenbereinigung folgt die Datenexploration, bei der die Daten visuell und statistisch untersucht werden. Ziel ist es, erste Muster und Zusammenhänge in den Daten zu erkennen.

Hierbei werden unter anderem Verteilungen, Korrelationen und Zeitreihenanalysen durchgeführt, um ein besseres Verständnis der Daten zu erlangen.

5. **Feature Engineering:** Im Feature Engineering werden neue Merkmale (Features) aus den vorhandenen Daten abgeleitet, die für die Analyse relevant sind.

Beispiele für solche Features sind:

- Durchschnittliche Geschwindigkeit: Berechnung der durchschnittlichen Geschwindigkeit für jede Fahrt.
- Beschleunigungsprofile: Analyse der Beschleunigungs- und Verzögerungsmuster.
- Energieverbrauch pro Kilometer: Berechnung des Energieverbrauchs pro zurückgelegtem Kilometer.

6. **Predictive Modeling:** In der Phase des Predictive Modeling werden statistische Modelle erstellt, um Vorhersagen und Analysen durchzuführen. Hierbei werden verschiedene Modellierungsansätze wie Regressionsanalyse, Varianzanalysen oder Hypothesentests verwendet, um signifikante Unterschiede und Zusammenhänge in den Daten zu identifizieren.
7. **Data Visualization:** Abschließend werden die Ergebnisse der Analyse visualisiert, um sie anschaulich darzustellen und leichter interpretieren zu können. Dies umfasst die Erstellung von Diagrammen, Grafiken und Karten, die die wesentlichen Erkenntnisse verdeutlichen. Die Visualisierung unterstützt dabei, die Ergebnisse effektiv zu kommunizieren und die Implikationen für die Nutzung von Elektrofahrzeugen im Alltag zu verstehen.

Durch die Anwendung des Data Science Lifecycles auf die Fahrdaten wird eine strukturierte und systematische Analyse ermöglicht, die fundierte und aussagekräftige Ergebnisse liefert.

Regression (wird komplett überarbeitet)

In diesem Kapitel wird eine detaillierte Analyse des Energieverbrauchs von Elektrofahrzeugen (EVs) und Verbrennungsmotorfahrzeugen (ICEs) durchgeführt. Dabei wird der Einfluss verschiedener Faktoren auf den Energieverbrauch untersucht, um ein besseres Verständnis der Effizienzunterschiede zwischen den beiden Fahrzeugtypen zu gewinnen.

Deskriptive Statistiken

Zunächst werden die deskriptiven Statistiken für den Energieverbrauch, die Durchschnittsgeschwindigkeit und die zurückgelegte Distanz für die beiden Fahrzeugtypen dargestellt. Die folgende Tabelle zeigt die mittleren Werte für diese Variablen:

Vehicle_Type	Energy_kWh_mean	Speed_kmh_mean	Distance_km_mean
<fct>	<dbl>	<dbl>	<dbl>
1 Golf	0.00154	28.3	0.00850
2 Mokka e	0.00152	31.1	0.00929

Listing 1: Deskriptive Statistiken

Visualisierung der Verteilung Energieverbrauchs

Die folgende Abbildung zeigt einen Boxplot des Energieverbrauchs für die beiden Fahrzeugtypen. Hierbei wird die Verteilung des Energieverbrauchs für den VW Golf 8 und den Opel Mokka E verglichen.

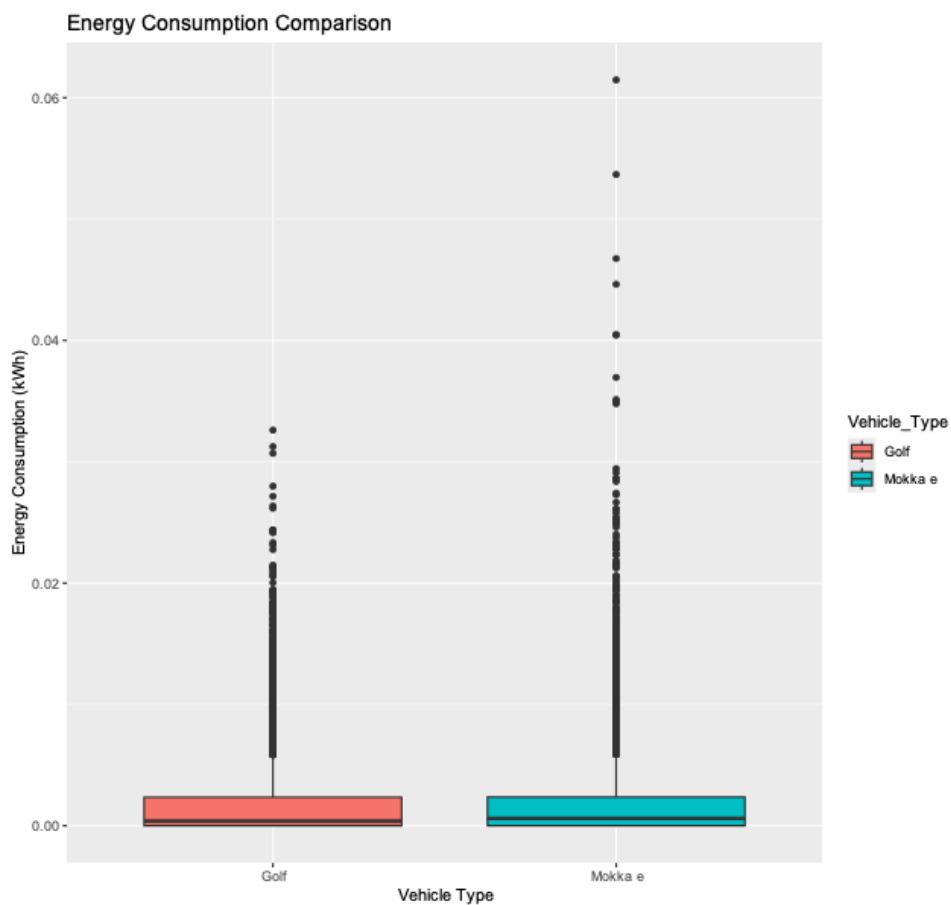


Figure 1: Vergleich des Energieverbrauchs beider Fahrzeugtypen

- Der mittlere Energieverbrauch ist bei beiden Fahrzeugtypen ähnlich, wobei der Opel Mokka E einen etwas niedrigeren Energieverbrauch aufweist.

- Die Verteilung der Energieverbrauchswerte zeigt, dass es beim VW Golf 8 mehr Ausreißer gibt, was auf eine größere Variabilität im Energieverbrauch hinweist.

Verteilung des Energieverbrauchs nach Geschwindigkeit

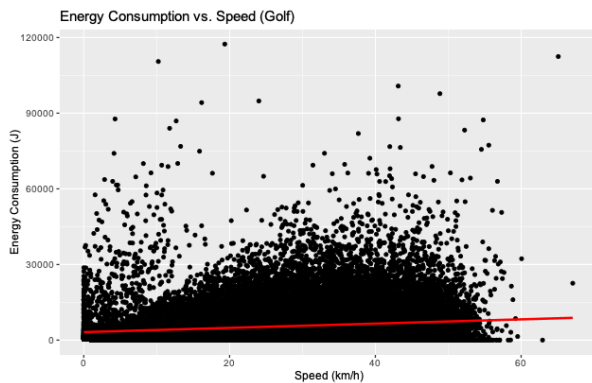


Figure 2: Verteilung Golf

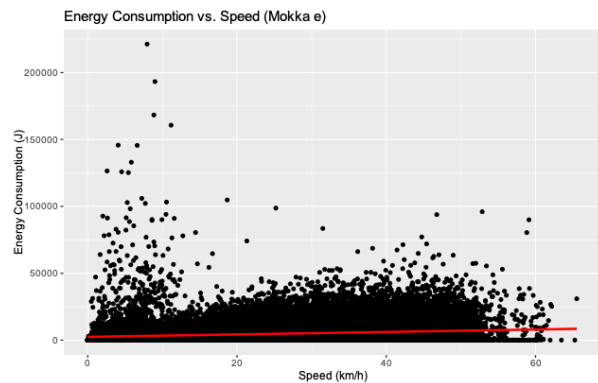


Figure 3: Verteilung Mokka

Correlation Funnel

für die Auswahl wichtiger Variablen für das Modell:

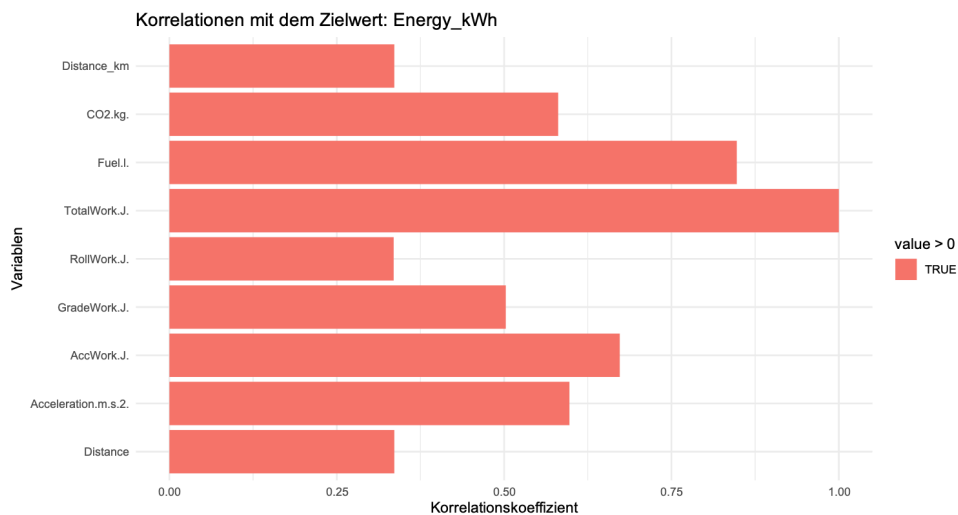


Figure 4: Verteilung Mokka

Regressionsanalyse

Um die Einflussfaktoren auf den Energieverbrauch genauer zu untersuchen, wurde eine lineare Regressionsanalyse durchgeführt. Die abhängige Variable ist der Energieverbrauch (kWh), während die unabhängigen Variablen der Fahrzeugtyp, die Geschwindigkeit, die Beschleunigung und die zurückgelegte Distanz sind.

```
energy_model <- lm(Energy_kWh ~ Vehicle_Type + Speed_kmh
+ Acceleration.m.s.2. + Distance_km, data = combined_data)
```

Listing 2: ModellFormel

```
Call:
lm(formula = Energy_kWh ~ Vehicle_Type + Speed_kmh + Acceleration.m.s.2. +
    Distance_km, data = combined_data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.029941	-0.000958	-0.000316	0.000650	0.047272

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.692e-04	1.248e-05	45.61	<2e-16 ***
Vehicle_TypeMokka e	-1.153e-04	1.142e-05	-10.10	<2e-16 ***
Speed_kmh	7.170e-06	3.416e-07	20.99	<2e-16 ***
Acceleration.m.s.2.	1.504e-03	6.411e-06	234.62	<2e-16 ***
Distance_km	8.979e-02	7.655e-04	117.29	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001688 on 99038 degrees of freedom

Multiple R-squared: 0.4563, Adjusted R-squared: 0.4563

F-statistic: 2.078e+04 on 4 and 99038 DF, p-value: < 2.2e-16

Listing 3: Modellformel und Ausgabe

- Der Intercept-Wert ist (5.692×10^{-4}), was der Grundverbrauch ohne Einfluss der erklärenden Variablen darstellt.
- Der Koeffizient für den Fahrzeugtyp (Mokka e) ist negativ und signifikant ($p < 2e-16$), was bedeutet, dass der Mokka e im Durchschnitt einen geringeren Energieverbrauch hat als der Golf.
- Die Geschwindigkeit ((Speed_kmh)) und Beschleunigung (Acceleration.m.s.2.) haben positive und signifikante Effekte auf den Energieverbrauch, was darauf hinweist, dass höhere Geschwindigkeiten und häufigere Beschleunigungen zu einem höheren Energieverbrauch führen.
- Die zurückgelegte Distanz (Distance_km) hat ebenfalls einen positiven und signifikanten Einfluss auf den Energieverbrauch.

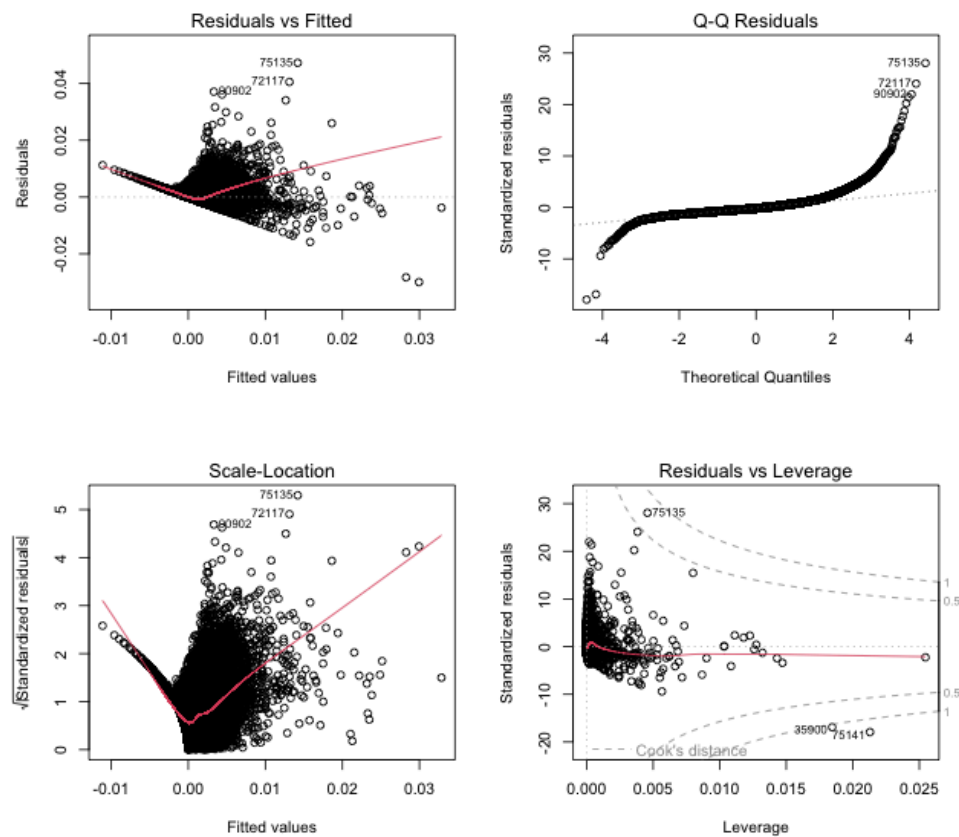


Figure 5: Diagnoseplots des Regressionsmodells

todo: zu regression Inhalt komplett überarbeiten

Random Forest (einbauen!)

Neuronales Netzwerk (einbauen)

Gradient Boosting Machines (GBM)

todo: einbauen

Schlussfolgerungen

Die Regressionsanalyse bestätigt, dass der Fahrzeugtyp, die Geschwindigkeit, die Beschleunigung und die zurückgelegte Distanz signifikante Einflussfaktoren auf den Energieverbrauch sind. Der Opel Mokka E zeigt im Durchschnitt einen geringeren Energieverbrauch als der VW Golf 8, was die Hypothese unterstützt, dass Elektrofahrzeuge effizienter im Energieverbrauch sind.

Vorschlag: Für zukünftige Analysen wäre es hilfreich, weitere erklärende Variablen zu berücksichtigen und möglicherweise nicht-lineare Modelle oder Interaktionseffekte zu untersuchen, um ein noch präziseres Modell des Energieverbrauchs zu entwickeln.

wltp-Ergebnisse

Die Ergebnisse der Analysen sind in den folgenden Abbildungen dargestellt. Die Balkendiagramme zeigen den Vergleich des Energieverbrauchs zwischen den realen Fahrbedingungen und den WLTP-Normwerten für beide Fahrzeugtypen und verschiedene Geschwindigkeitskategorien.

VW Golf 8 2.0 TDI

- Extra High Speed: real 4.5 kWh/100km, WLTP 4.5 kWh/100km

- High Speed: real 3.7 kWh/100km, WLTP 3.7 kWh/100km
- Low Speed: real 6.58 kWh/100km, WLTP 6.2 kWh/100km
- Medium Speed: real 9.87 kWh/100km, WLTP 4.3 kWh/100km

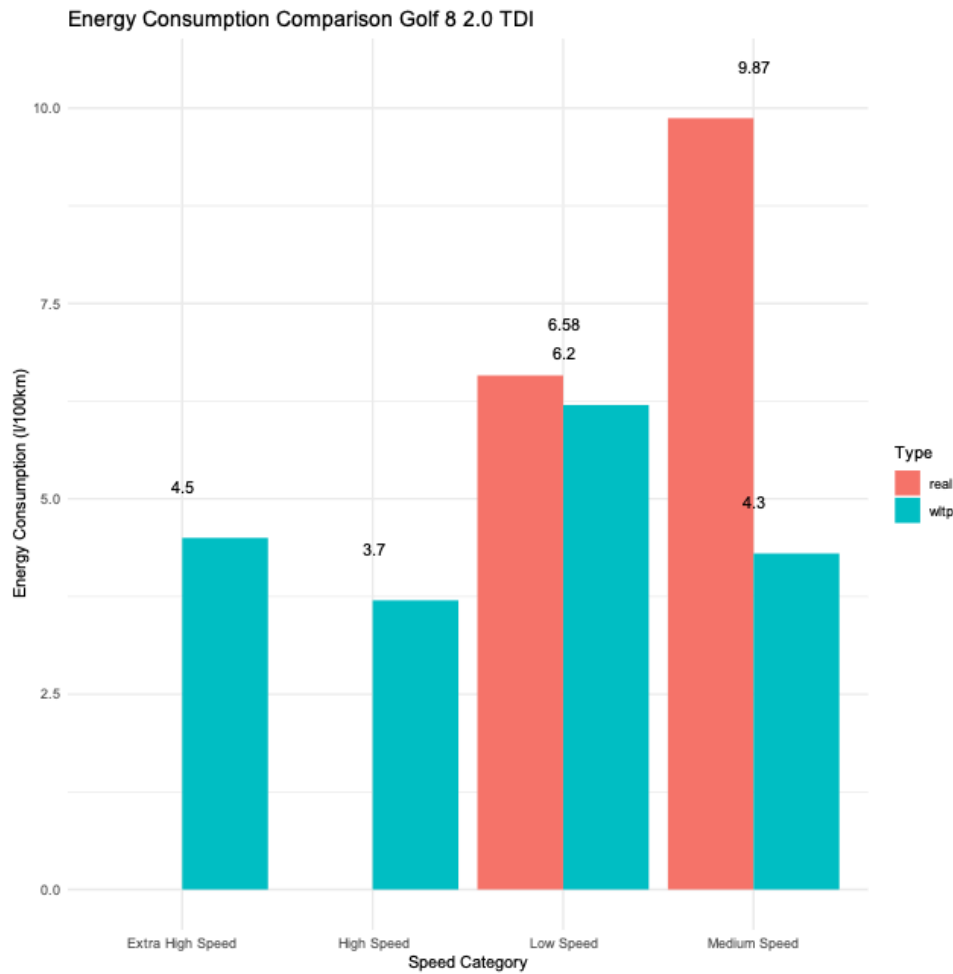


Figure 6: WLTP Vergleich Golf

Opel Mokka e

- Extra High Speed: real 20.22 kWh/100km, WLTP 20.22 kWh/100km
- High Speed: real 17.8 kWh/100km, WLTP 17.8 kWh/100km
- Low Speed: real 17.16 kWh/100km, WLTP 15.8 kWh/100km
- Medium Speed: real 21.35 kWh/100km, WLTP 16.2 kWh/100km

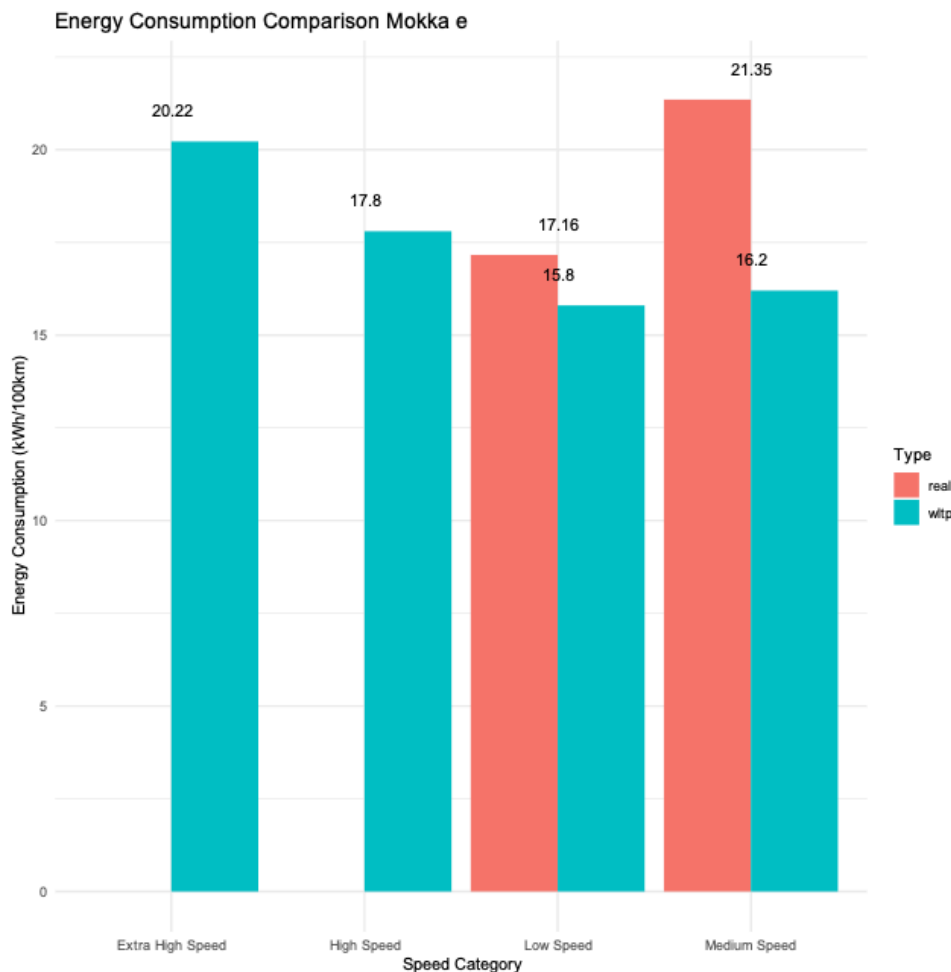


Figure 7: Diagnoseplots des Regressionsmodells

Diskussion der Ergebnisse

Signifikante Diskrepanzen

Die Analyse zeigt signifikante Diskrepanzen zwischen den realen Energieverbrauchswerten und den WLTP-Normwerten, insbesondere bei niedrigen und mittleren Geschwindigkeiten. Diese Diskrepanzen deuten darauf hin, dass die WLTP-Werte nicht immer die realen Fahrbedingungen adäquat widerspiegeln.

VW Golf 8 2.0 TDI

- Niedrige und mittlere Geschwindigkeiten: Die realen Verbrauchswerte zeigen eine signifikant höhere Abweichung von den WLTP-Werten, was auf Ineffizienzen des Fahrzeugs bei unterschiedlichen Fahrbedingungen hinweist. Dies könnte auf Faktoren wie variierende Fahrstile, Verkehrsbedingungen und Umweltfaktoren zurückzuführen sein .
- Hohe und sehr hohe Geschwindigkeiten: Die realen und WLTP-Werte stimmen weitgehend überein, was auf eine bessere Leistung des Fahrzeugs unter diesen Bedingungen hindeutet .

Opel Mokka e

- Niedrige und mittlere Geschwindigkeiten: Auch hier sind die Diskrepanzen deutlich ausgeprägt, wobei die realen Verbrauchswerte höher sind als die WLTP-Werte. Dies könnte auf die höhere Empfindlichkeit von Elektrofahrzeugen gegenüber variablen Fahrbedingungen hinweisen, insbesondere bei niedrigen Geschwindigkeiten und häufigen Beschleunigungs- und Bremsmanövern .

- Hohe und sehr hohe Geschwindigkeiten: Der Unterschied zwischen den realen und WLTP-Werten ist weniger signifikant, was auf eine effizientere Nutzung der Energie bei höheren Geschwindigkeiten hinweist .

Analysemethodik

Die Analyse der erhobenen Fahrdaten erfolgt in zwei Schwerpunkten, die den Hypothesen H1 und H2 entsprechen. Für H1 steht die Untersuchung der Energie- und Beschleunigungsperformance im Vordergrund, während für H2 die Einteilung der Fahrdaten in WLTP-Kategorien und die Analyse der Abweichungen zwischen Soll- und Realverbrauch relevant sind.