



BAKERY SALES PREDICTION GRUPPE 14

Von Stefan Linger und Lasse Schettlinger

Warengruppen: Dummy Variablen für die Warengruppen

Kieler Woche: Dummy Variable KiWo

Feiertage: Dummy Variable für Feiertage in SH

Regenwetter: Dummy Variable für Regen

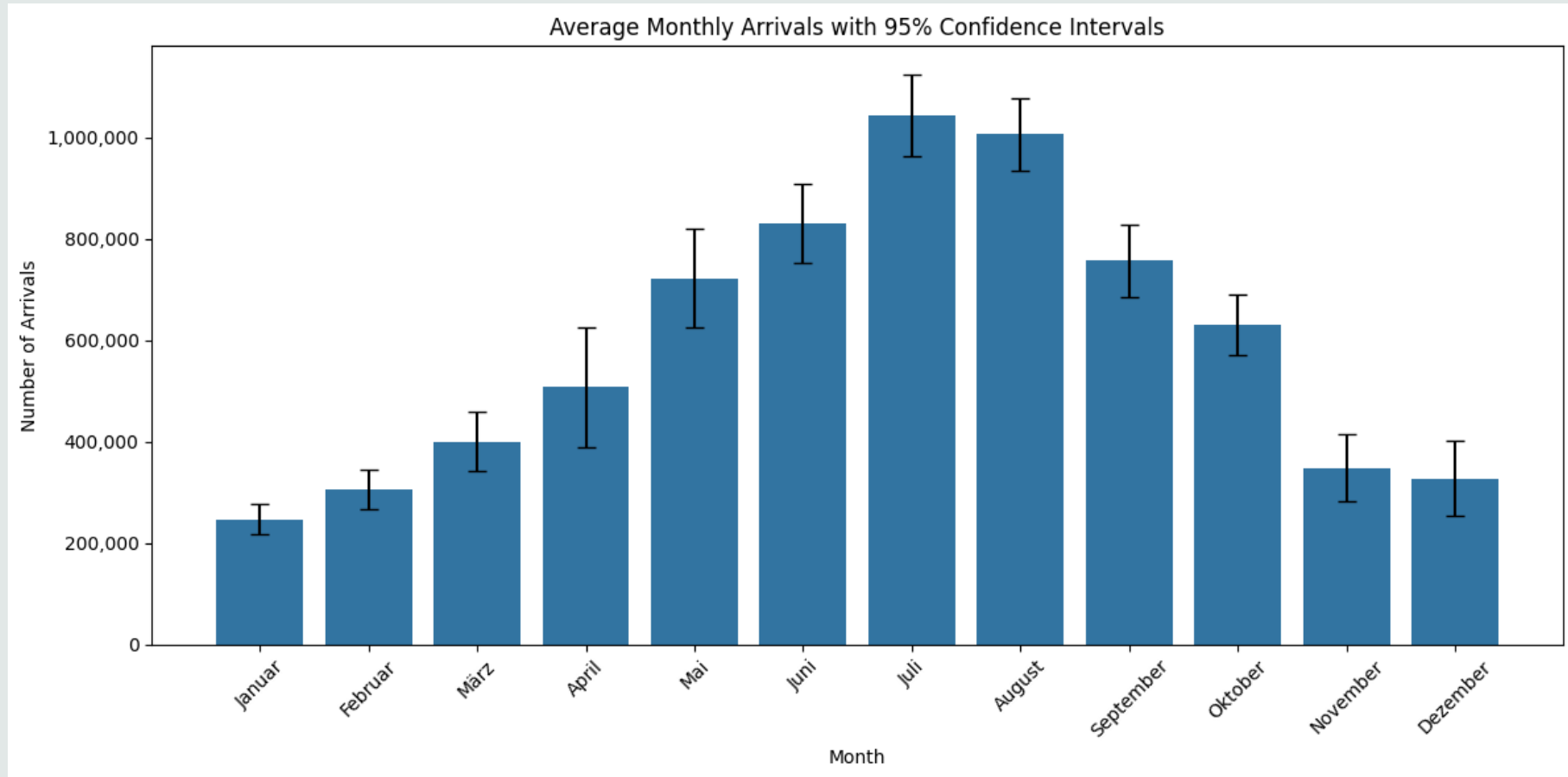
Schneewetter: Dummy Variable für Schnee

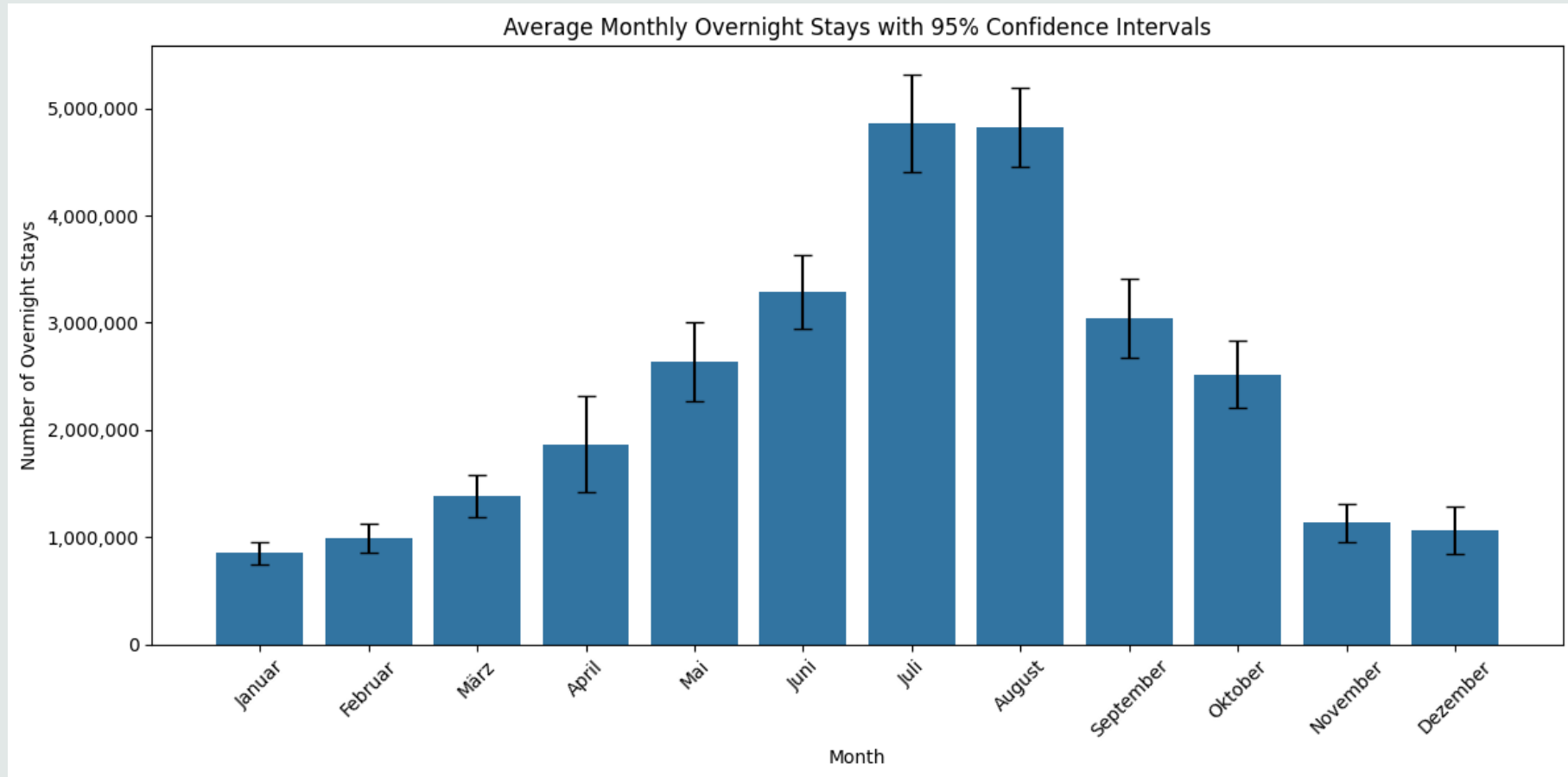
Wochentage: Dummy Variablen für die Tage der Woche

Temperatur: Dummy Variablen für die eine hohe, mittlere oder niedrige Temperatur gruppiert nach Jahreszeit

Tourismus: monatliche Übernachtungen und Ankünfte für Tourismus in SH

Features





Modellgleichung für die OLS-Regression:

Umsatz ~ snow_weather + rain_weather +
Windgeschwindigkeit + KielerWoche + Feiertag +
is_weekend + Season_Temp_Fall_high +
Season_Temp_Fall_low + Season_Temp_Spring_high +
Season_Temp_Spring_middle + Season_Temp_Spring_low +
Season_Temp_Summer_high + Season_Temp_Summer_low
+ Season_Temp_Summer_middle +
Season_Temp_Winter_high + Season_Temp_Winter_middle
+ Season_Temp_Winter_low + Warengruppe_1 +
Warengruppe_2 + Warengruppe_3 + Warengruppe_6 +
Warengruppe_5

Adjusted R²:

0.78623527443317 -> ca. 78.62%

Linear Model

Missing Value Imputation:

KNN Imputation

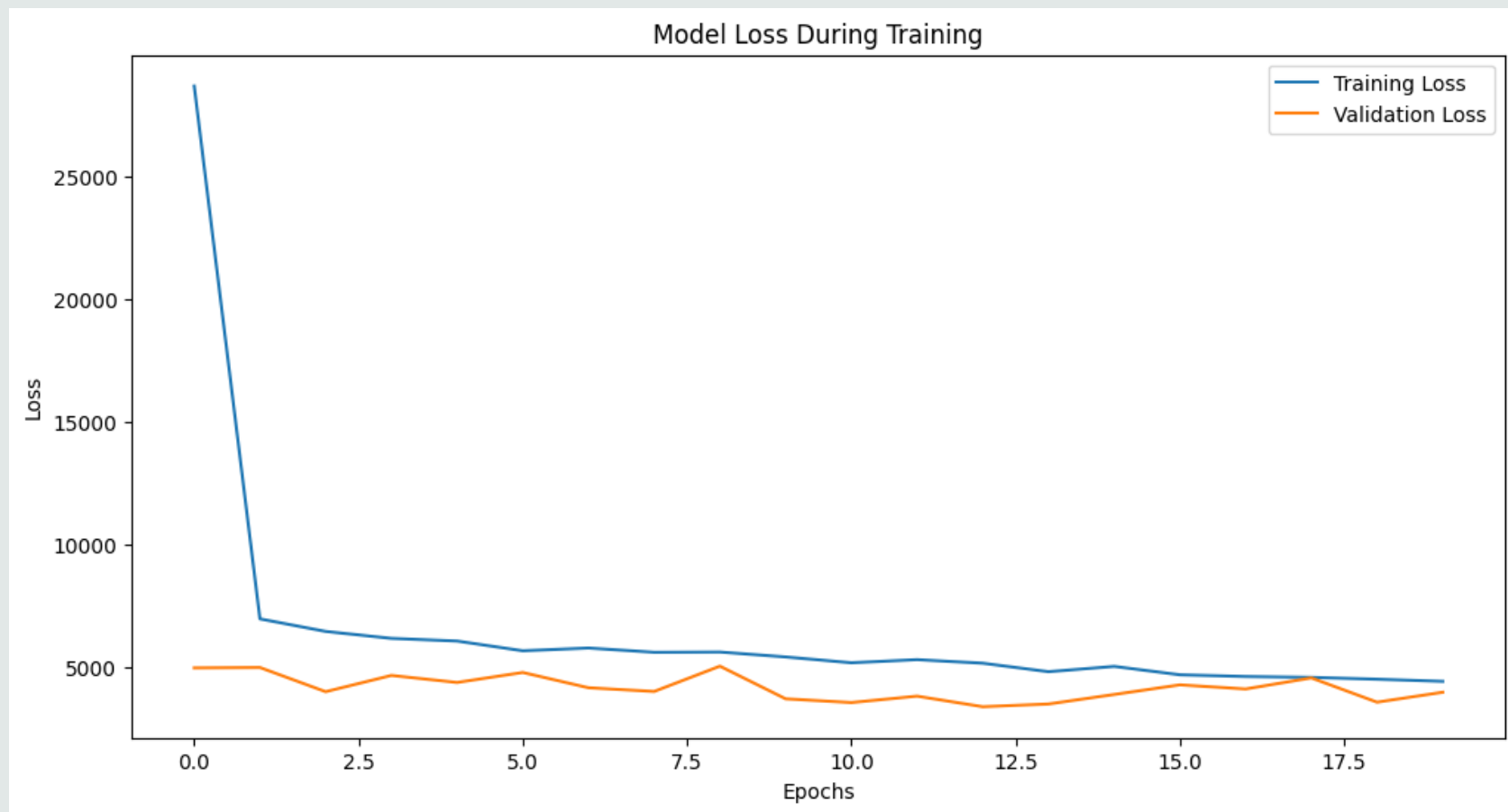
**Fehlende
Werte**

Neuronales Netzwerk

```
model = Sequential([
    InputLayer(shape=(training_features.shape[1], )),
    BatchNormalization(),
    Dense(100, activation='relu'),
    Dense(40, activation='relu'),
    Dense(10, activation='relu'),
    Dense(1)
])

model.compile(loss="mse", optimizer=Adam(learning_rate=0.001))

history = model.fit([training_features, training_labels, epochs=20,
                    validation_data=(validation_features, validation_labels)])
```



Mean Absolute Percentage Error:

MAPE auf Training Data: 0.20

MAPE auf Validation Data: 0.24

MAPE auf Warengruppe 1: 0.21

MAPE auf Warengruppe 2: 0.23

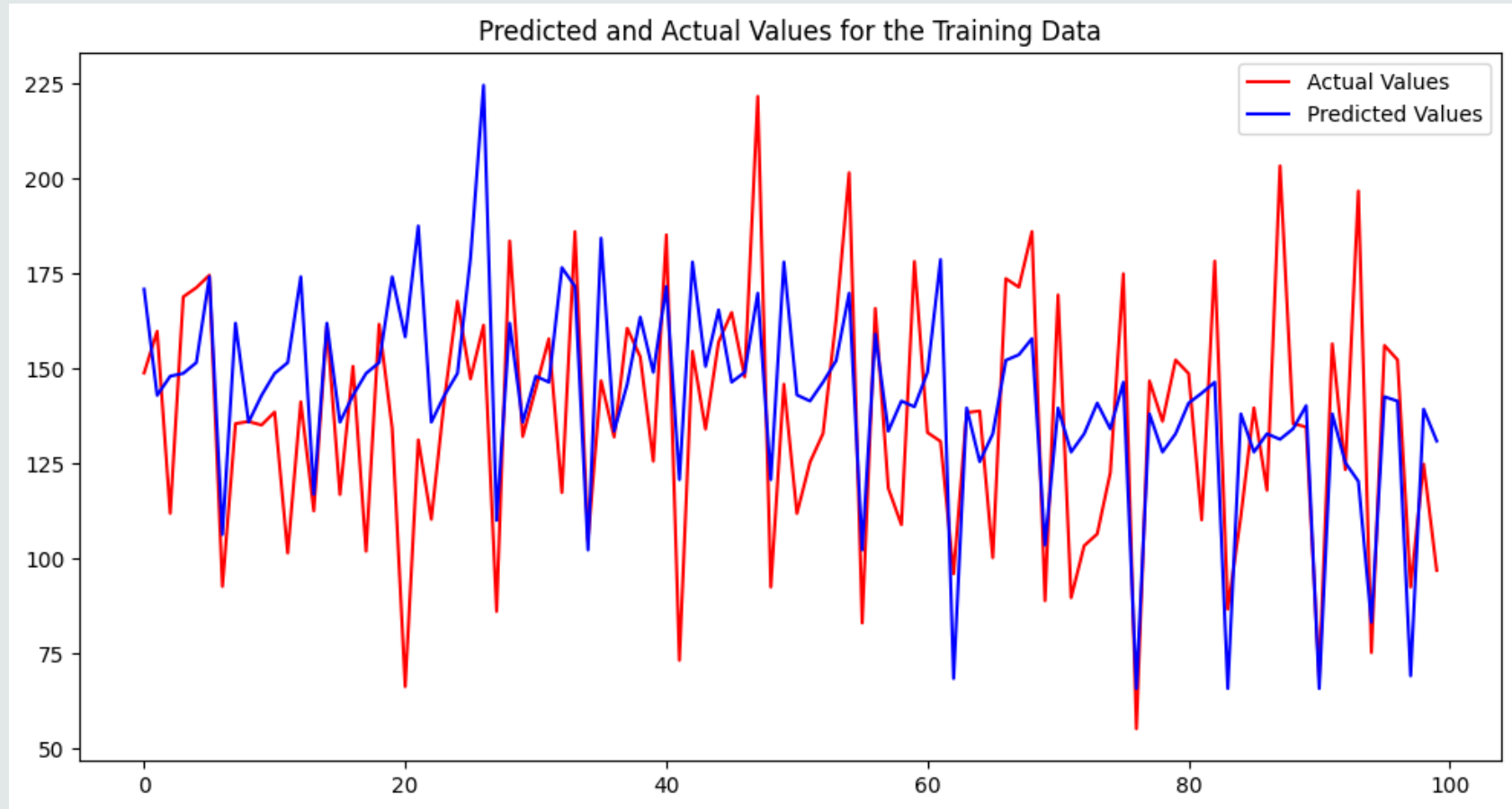
MAPE auf Warengruppe 3: 0.25

MAPE auf Warengruppe 4: 0.27

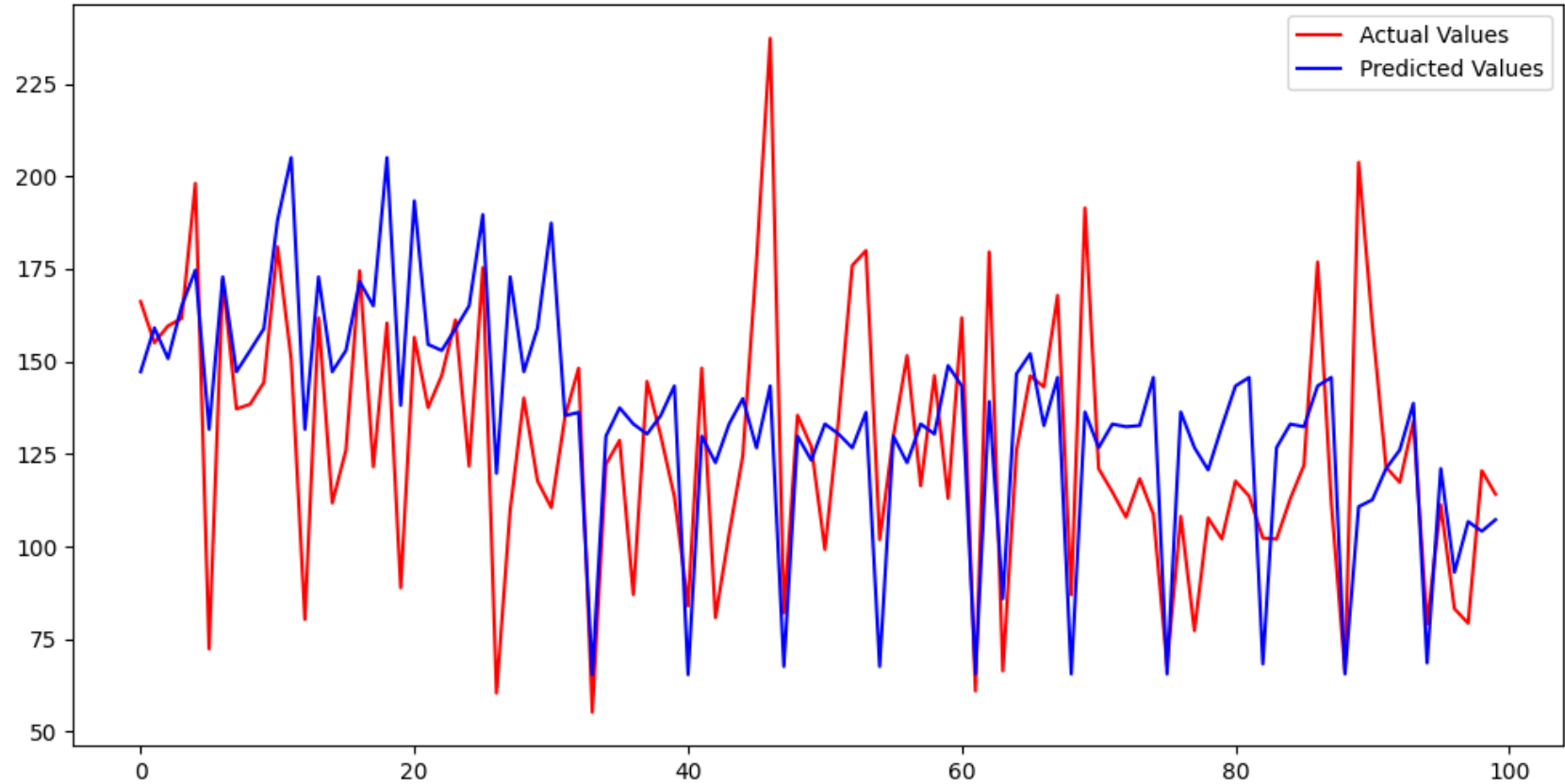
MAPE auf Warengruppe 5: 0.20

MAPE auf Warengruppe 6: 0.55

MAPE



Predicted and Actual Values for the Validation Data



Wir hatten versucht einen Lag für den Umsatz in das Modell einzubinden, was auf den Trainingsdaten, schon mit der Regression sehr gute Ergebnisse erzeugt hat.

Nur kann so natürlich keine Voraussage stattfinden, wenn wir keine Zeitreihenanalyse machen, da das Modell (Regression und unser neurales Netz) den Umsatz von vor einer Woche nicht miteinbeziehen kann, da unbekannt.

Daraus ergibt sich die Überlegung bspw. auch ein anderes Modell für die Voraussage zu benutzen, wenn man es noch verbessern will.

**„Worst
Fail“**