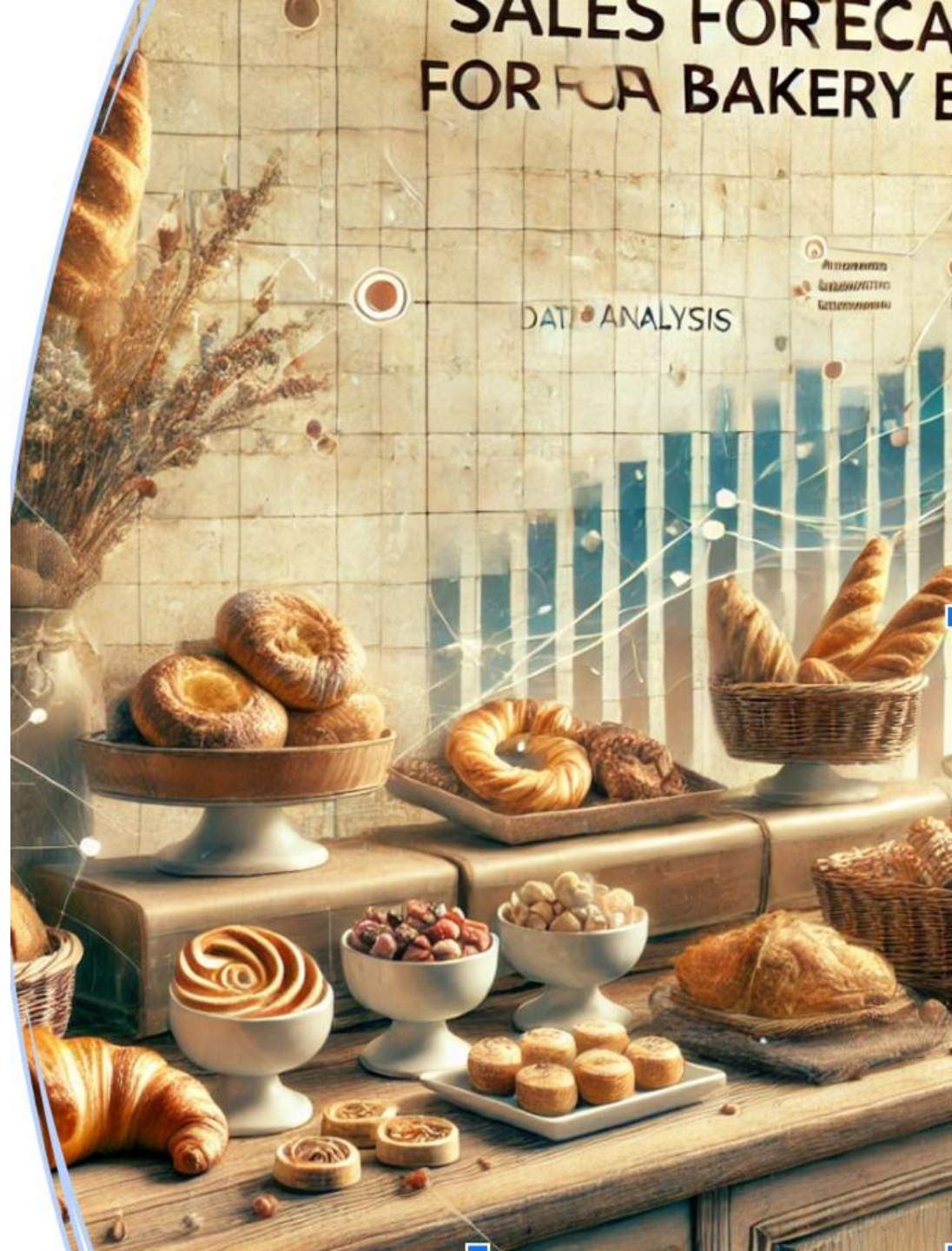


Umsatzvorhersage einer Bäckereifiliale



Vorhersage
zukünftiger Umsätze
für sechs
Bäckereiproduktkategorien
anhand
historischer Daten von
2013 bis 2018.

Team 08

Teammitglieder:

- **Dirk Brockhausen:**
Datensatzaufbereitung,
Baseline-Modell,
Modellentwicklung,
Visualisierung,
Ergebniszusammenstellung

- **Miriam Berger:**
Datensatzaufbereitung,
Baseline-Modell, Team-
Koordination,
Dokumentation

Stand: (16.01.2025)



Variablen

Die durch die Competition Data vorgegebenen Variablen sind:

- Warengruppe
- Bewölkung
- Temperatur
- Windgeschwindigkeit
- Wettercode
- Kieler Woche

Die Zielvariable ist:

- Umsatz

Neue Variablen nach Kategorien

Zeitliche Faktoren:

- Silvester, Sommer, Wochenenden, Oster-Samstag, Wochentag, Tag vor Feiertagen

Temperaturmerkmale:

- Gefühlte Temperatur, Temperaturkategorien (warm, kalt, mild)

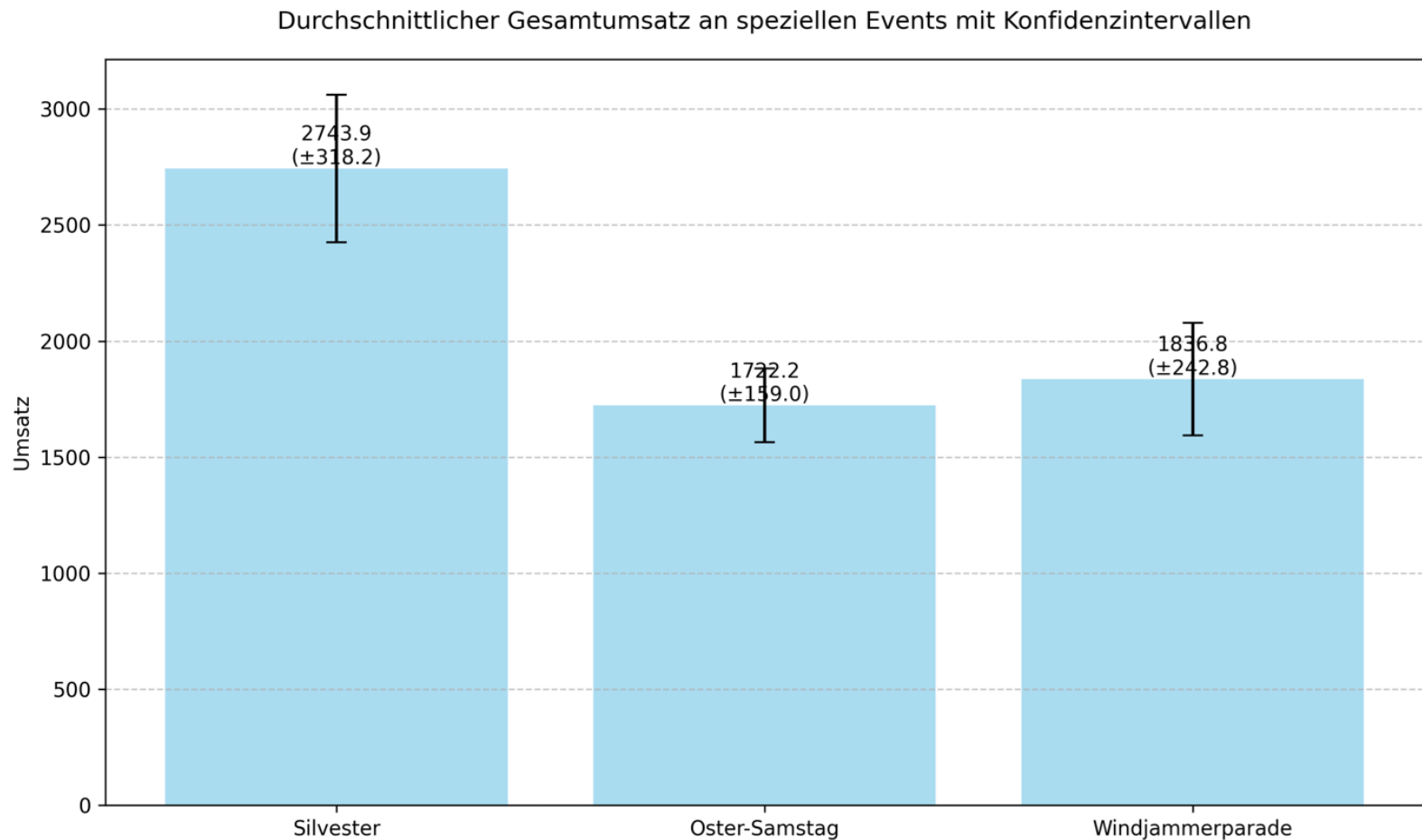
Schulferien & Events:

- Schulferien, Feiertage

Wettermerkmale:

- Jahreszeiten, Niederschlagsereignisse (Schnee, Nebel)

Balkendiagramme für drei selbst erstellte Variablen



Optimierung des linearen Modells

Training Metrics:

RMSE: 72.50, MAE: 49.90, R^2 : 0.759, Adjusted R^2 : 0.758

Validation Metrics:

RMSE: 69.73, MAE: 51.43, R^2 : 0.713, Adjusted R^2 : 0.708

Model Equation:

Sales = β_0 +

Very High Importance Features

$\beta_1(\text{is_silvester}) + \beta_2(\text{is_summer}) +$
 $\beta_3(\text{temp_base_warm}) +$
 $\beta_4(\text{is_weekend}) + \beta_5(\text{is_good_weather}) +$
 $\beta_6(\text{product_group}) +$
 $\sum_i \beta_i(\text{school_holiday_state}_i) +$

High Importance Features

$\beta_7(\text{day_of_week}) + \beta_8(\text{temperature}) +$
 $\beta_9(\text{feels_like_temp}) +$
 $\beta_{10}(\text{cloud_cover}) + \beta_{11}(\text{year}) + \beta_{12}(\text{quarter}) +$

Moderate Importance Features

$\beta_{13}(\text{is_winter}) + \beta_{14}(\text{temp_base_cold}) +$
 $\beta_{15}(\text{temp_base_mild}) +$
 $\beta_{16}(\text{is_day_before_holiday}) +$
 $\beta_{17}(\text{is_easter_saturday}) +$
 $\beta_{18}(\text{is_month_end}) + \beta_{19}(\text{is_kieler_woche}) +$

Lower Importance Features

$\beta_{20}(\text{is_spring}) + \beta_{21}(\text{is_fall}) + \beta_{22}(\text{is_public_holiday})$
+
 $\beta_{23}(\text{is_windjammer}) + \sum_i \beta_i(\text{weather_condition}_i)$

Missing Value Imputation (Neural Network v14)

Verschiedene Imputationsmethoden nach Datentyp

Numerische Variablen

Verwendung des Mittelwerts für Wetterdaten

Mean Imputation für Temperatur

Kategorische Variablen

Logischer Defaultwert für fehlende Werte

Konstante Imputation

Fehlender Wettercode mit -1 aufgefüllt

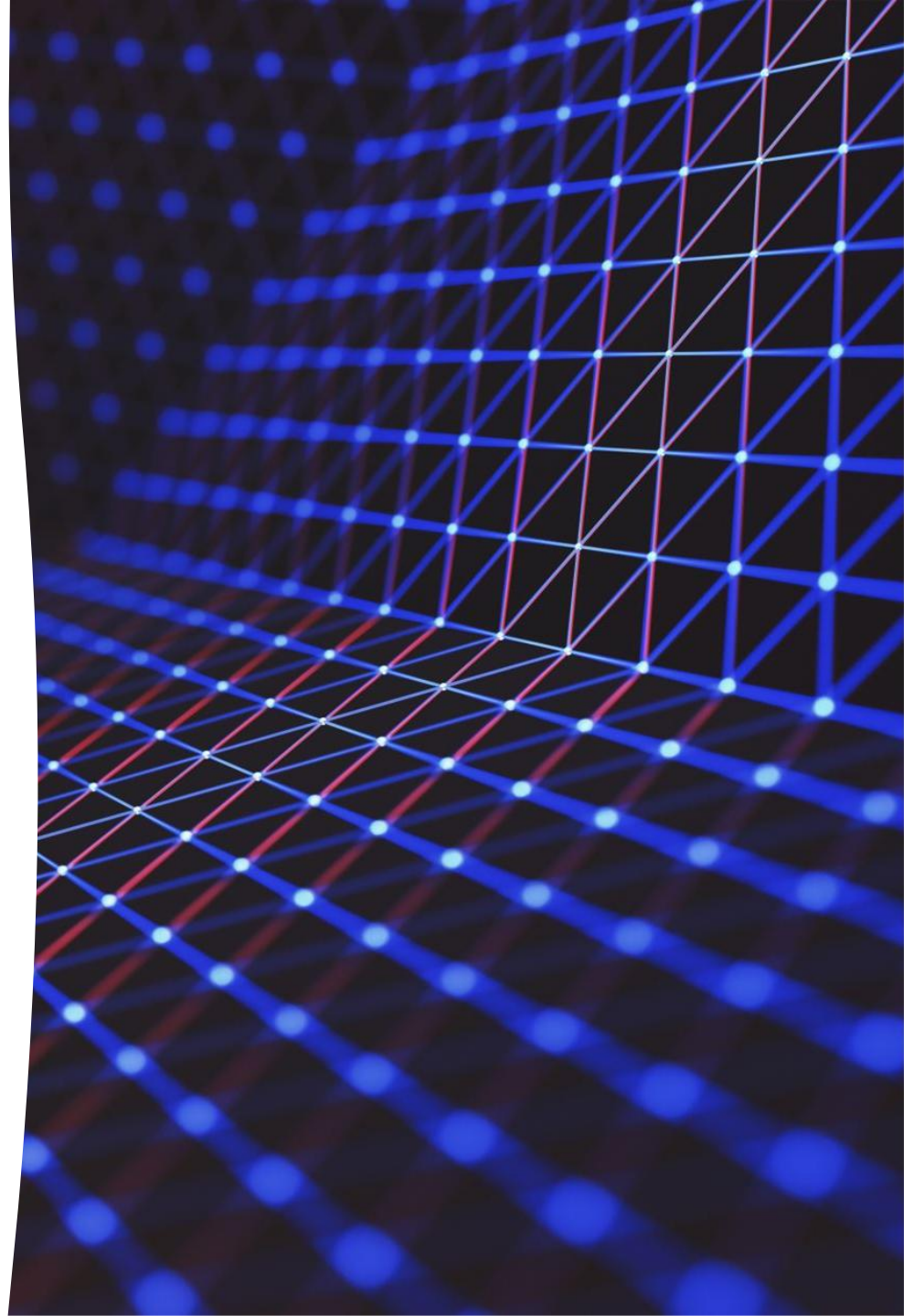
Binäre Imputation

Fehlende Events (Kieler Woche, Ferien, etc.) mit 0 aufgefüllt, was kein “Event” bedeutet.

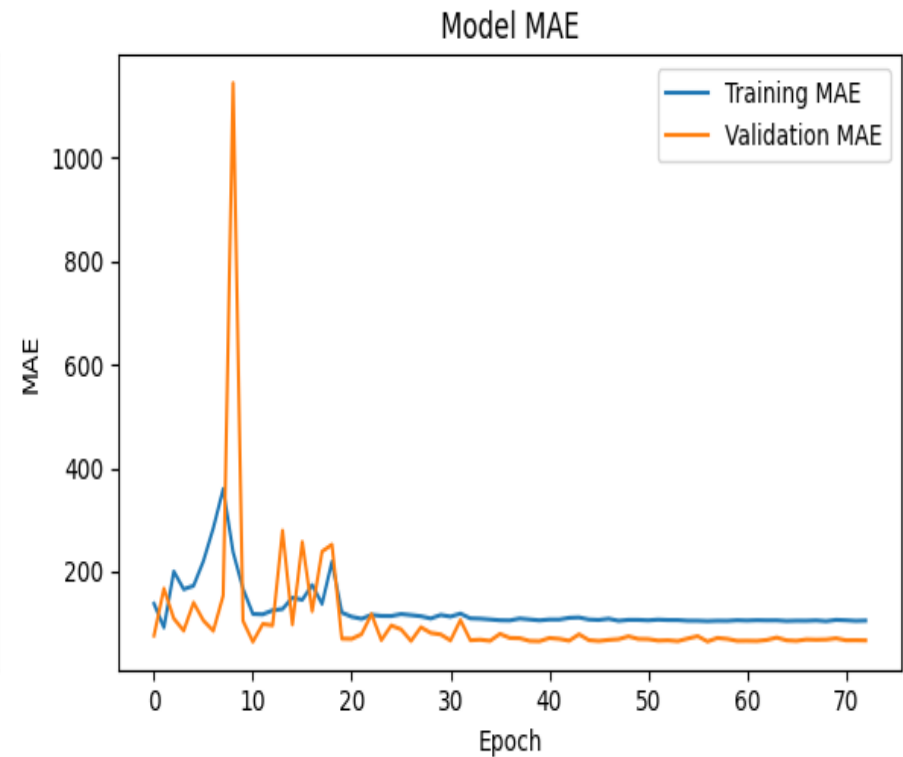
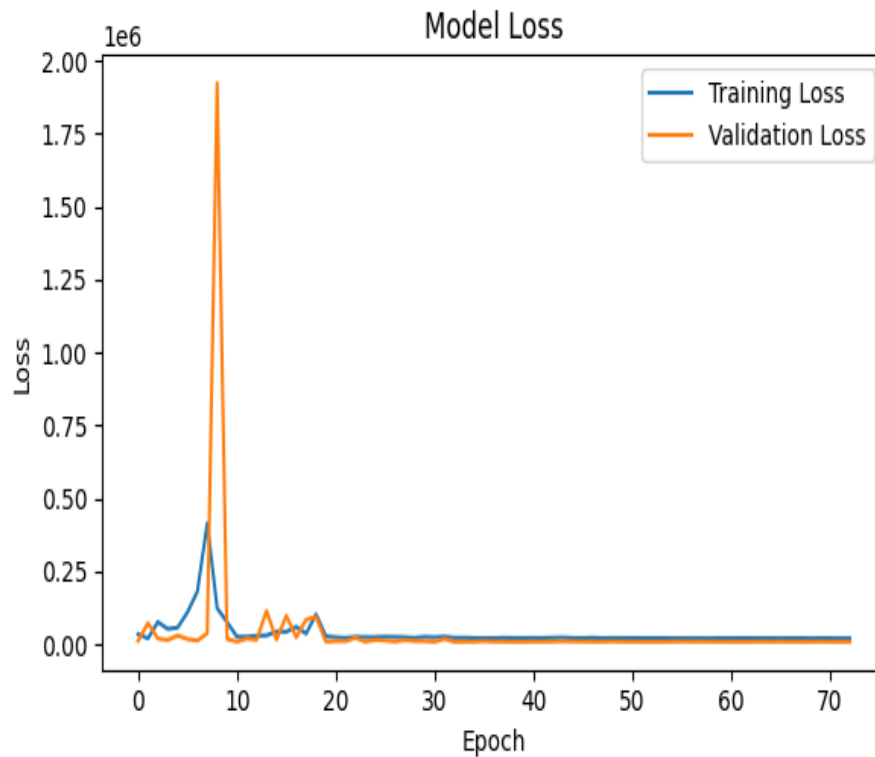
Optimization of the Neural Network

Source code for defining the neural network:

```
def create_neural_network(input_shape):  
    model = keras.Sequential([  
        layers.InputLayer(input_shape=(input_shape,)),  
        layers.BatchNormalization(),  
        layers.Dense(128, activation='relu'),  
        layers.Dropout(0.3),  
        layers.Dense(64, activation='relu'),  
        layers.Dropout(0.3),  
        layers.Dense(32, activation='relu'),  
        layers.Dense(1)  
    ])
```



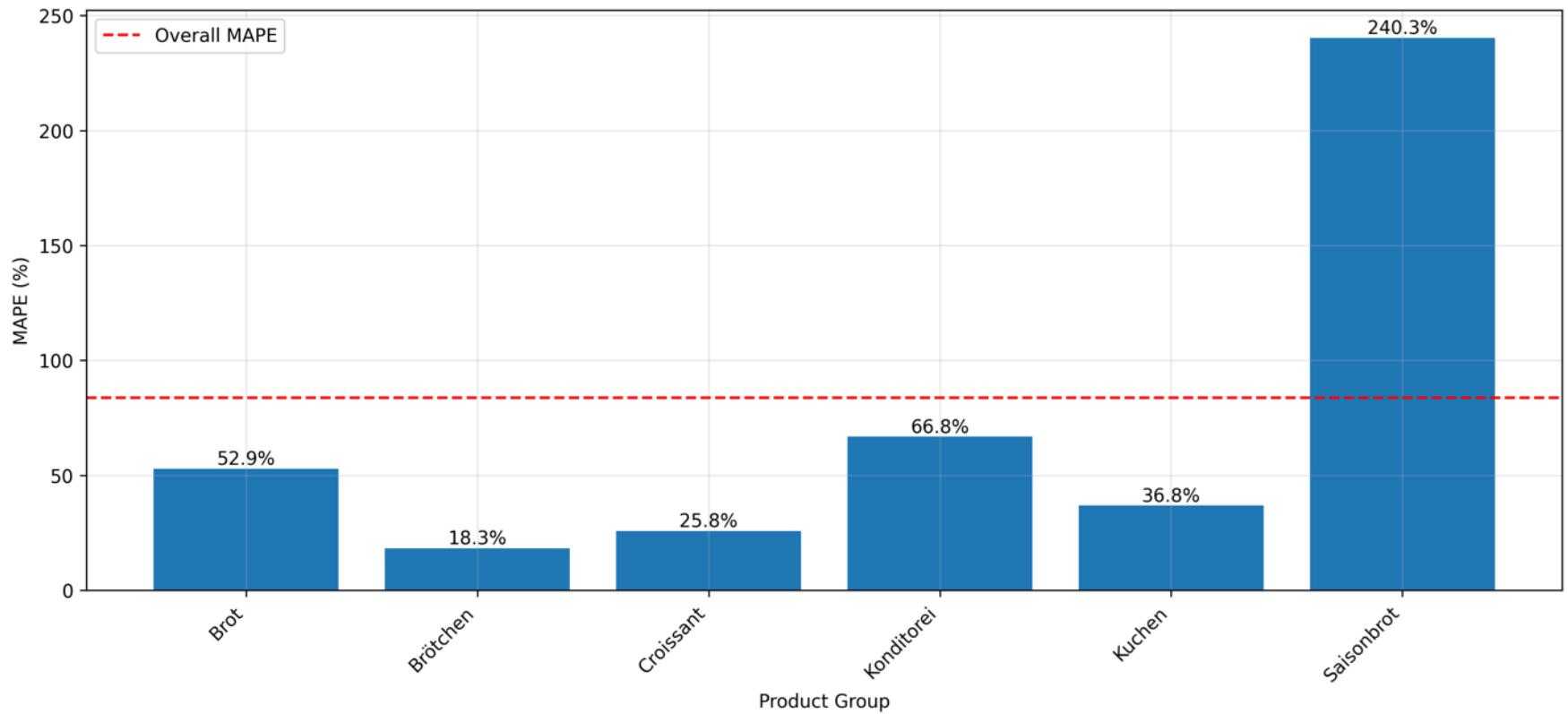
Visualization of loss functions for training and validation datasets



MAPEs

MAPEs für Validierungsdatensatz und Warengruppen

MAPE by Product Group



Worst Fails

Probleme:

- Zu viele Variablen
 - > Reduktion von 66 auf 23 (10 Basis-Variablen + 7 Wochentagsvariablen + 6 Warengruppen-Variablen)
- Unzureichende KI-Entwicklungsumgebungen
(Cursor und Windsurf)

Vielen Dank!

Vielen Dank für
eure
Aufmerksamkeit
!

Habt ihr Fragen?