Bakery Sales Prediction

Benedikt Luther

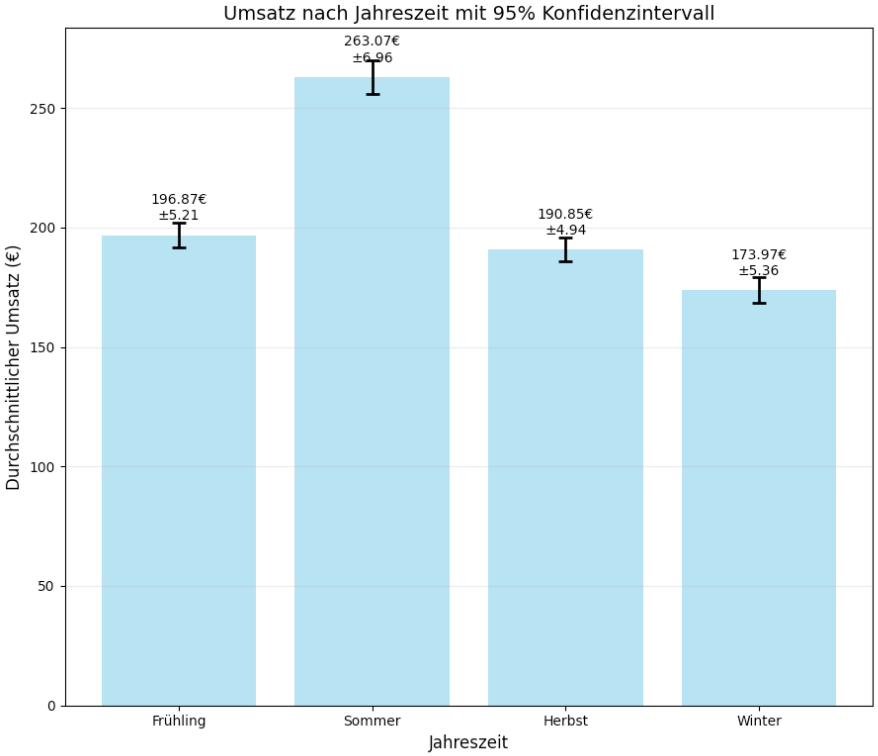
Natnael Guesh Hagos,

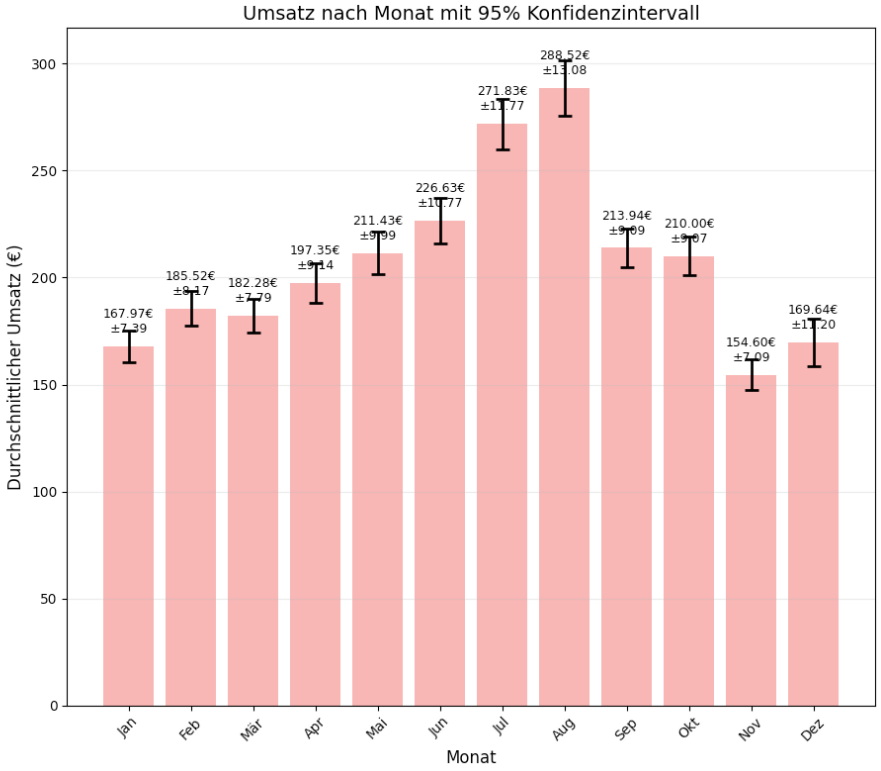
Variable / Group	English Translation	Encoding / Type
Jahr	Year	Integer
Monat	Month	Integer (1–12)
Tag	Day	Integer (1–31)
Wochentag	Weekday Name	String
Wochentag_Nr, Wochentag_Nr.1	Weekday Number	Integer (0–6)
Warengruppe	Product Group ID	Integer
Wettercode_fehlt	Weather Code Missing	Binary (0/1)
ist_feiertag	Is Holiday	Binary (0/1)
ist_kiwo	Is Kiel Week Event	Binary (0/1)
Warengruppe_Brot, Warengruppe_Brötchen, Warengruppe_Croissant,Warengruppe_Konditorei, Warengruppe_Kuchen, Warengruppe_Saisonbrot	Product Group Dummies:Bread, Rolls, Croissant, Pastry Shop, Cake, Seasonal Bread	One-hot encoded (Binary 0/1)
Jahreszeit	Season (e.g., Summer)	String
Jahreszeit_Winter, Jahreszeit_Frühling, Jahreszeit_Sommer, Jahreszeit_Herbst	Season Dummies:Winter, Spring, Summer, Autumn	One-hot encoded (Binary 0/1)
Wochentag_Monday, Wochentag_Tuesday, Wochentag_Wednesday,Wochentag_Thursday, Wochentag_Friday, Wochentag_Saturday, Wochentag_Sunday	Weekday Dummies:Monday to Sunday	One-hot encoded (Binary 0/1)
Preis Index	Pricing Index / Inflation Rate	

Missing Value Imputation

Variable	Suggested Type	Imputation Method	
Temperatur	Missing at Random	Mean Imputation	Simple average used
Windgeschwindig-keit	Missing at Random	Mean Imputation	Rare missing values
Bewoelkung	Missing at Random	Mean Imputation	Few gaps, imputed globally
Wettercode	Systematic / Not Random	External API Request	Retrieved from weather API if missing
Wettercode_fehlt			Indicates if weather code was missing (0/1)

Revenue (Seasons and Months)





Some explorative Insights

```
=== 6.3 WETTERBEDINGTE VERKAUFSMUSTER ===
📊 Umsatz nach Wetter-Kategorien:
```

Wetter_Kategorie	mean	count	std	CV
Durchschnitt	188.55	3326	128.47	0.681
Kalt/Bewölkt	169.54	1740	118.34	0.698
Mild & Heiter	239.17	2125	154.38	0.645
Sonnig & Warm	258.84	1404	175.29	0.677
Windig	184.10	739	128.10	0.696

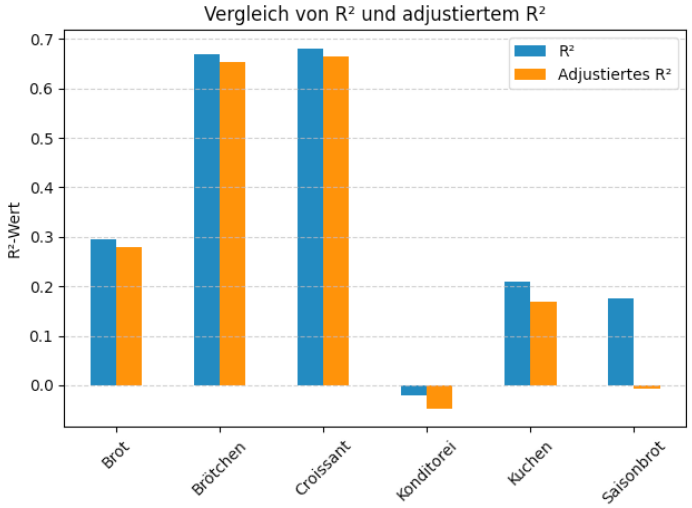
Key Insights:

- Best weather for revenue: **Sunny & warm**
- Worst weather for revenue: **Cold/cloudy**
- Most weather-sensitive product group: **Rolls**
- Most weather-stable product group: **Pastry shop**

Improved (Linear) Model

In comparison to the baseline

- Best Win: Regression for each product group
- Worst Fail: Including the inflation data



Improved (Linear) Model

Croissant RidgeCV

Umsatz = 76.85 -36.82*ist_kiwo +29.04*ist_feiertag -27.49*Wochentag_Friday -21.94*Wochentag_Thursday +19.31*Jahreszeit_Sommer + ... R^2: 0.6816 Adjustiertes R^2: 0.6646 Ausgewählte Features: 18/21

Brötchen RidgeCV

Umsatz = 201.77 +291.23*ist_feiertag +114.53*Jahreszeit_Sommer -81.83*Wochentag_Friday +55.60*Wochentag_Monday -54.51*Wochentag_Thursday + ... R^2: 0.6688 Adjustiertes R^2: 0.6533 Ausgewählte Features: 16/21

...

	MAE	R^2	Adjustiertes R^2	Modell-Typ	Ausgewählte Features
Brot	23.404287	0.295116	0.278911	GradientBoostingRegressor	8
Brötchen	58.194677	0.668840	0.653256	RidgeCV	16
Croissant	31.552284	0.681563	0.664605	RidgeCV	18
Konditorei	20.215126	-0.020782	-0.047258	RidgeCV	9
Kuchen	44.586274	0.209975	0.167903	RidgeCV	18
Saisonbrot	20.464855	0.176579	-0.006403	GradientBoostingRegressor	10

MAPE: 0.20495

Neural Net (Work in Progress)

```python
# 🧠 Iteration 2: Erweiterte Modell-Architektur
def create_extended_neural_network(input_dim):
    """Erstellt ein größeres neuronales Netzwerk mit Dropout für die erweiterten Features"""

    model = Sequential([
        Dense(128, activation='relu', input_shape=(input_dim,)),
        Dropout(0.2),
        Dense(64, activation='relu'),
        Dropout(0.2),
        Dense(32, activation='relu'),
        Dense(1, activation='linear')  # Linear für Regression
    ])

    # Kompiliere Modell
    model.compile(
        optimizer='adam',
        loss='mse',
        metrics=['mae']
    )

    return model
```

Neural Net (Work in Progress)

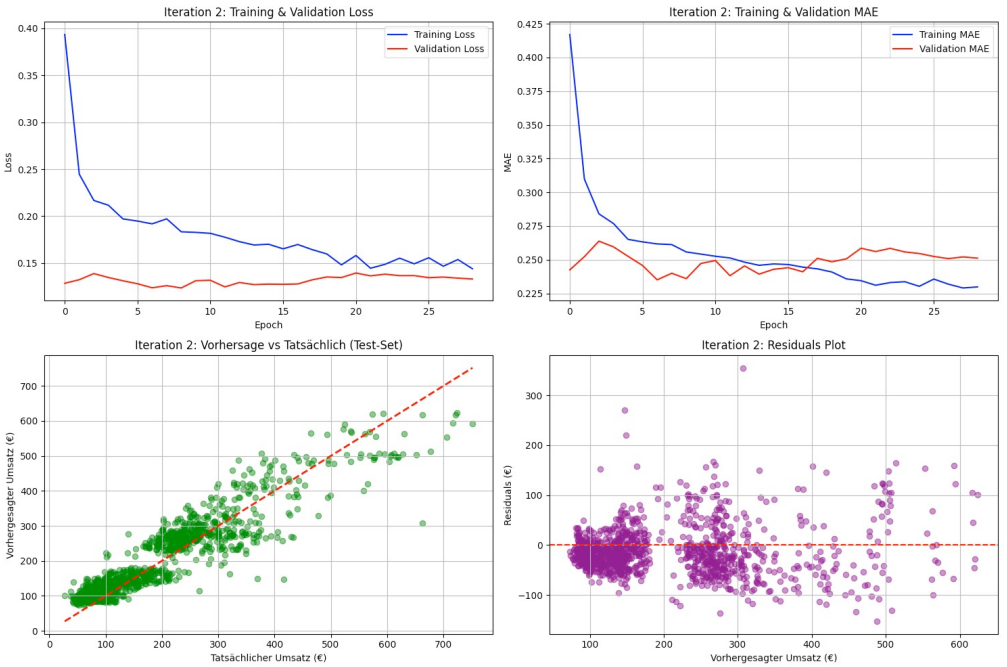
MAPE: 0.20563

MODEL ARCHITECTURE:

Iteration 1: Dense(64) → Dense(32) → Dense(1)
Iteration 2: Dense(128) → Dropout(0.2) → Dense(64) → Dropout(0.2) → Dense(32) → Dense(1)

FEATURE ENGINEERING:
Iteration 1: 17 features
Iteration 2: 134 features

MODEL PARAMETERS:
Iteration 1: 3,265 parameters
Iteration 2: 27,649 parameters

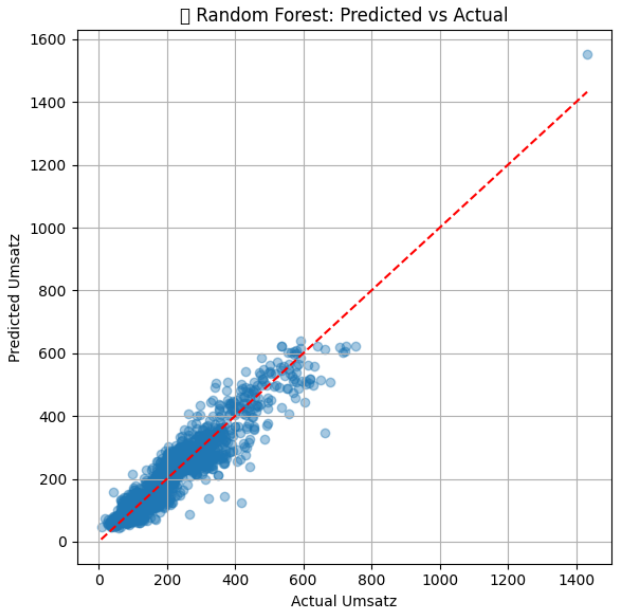


PERFORMANCE COMPARISON (Test Set):
Iteration 1 test results not available
Iteration 2 - MAE: €38.77, RMSE: €51.53

Random forest

Random Forest
Evaluation Results

- MAE: 29.65
- RMSE: 41.92
- R²: 0.896



Additional Steps for our Project

- Improvement of the neural net (work in progress) Normalization etc.
- Adding even more data
- Trying other imputations
- Using other methods to reduce outlier

Thank you!