

# Vorhersage des Umsatzes einer Kieler Bäckerei

**Opencampus Kurs:**  
Einführung in Data Science  
und maschinelles Lernen mit R

**Gruppe 3:**  
Niels, Kathrin, Svetlana, Nora

# Formatierung des Datensatzes

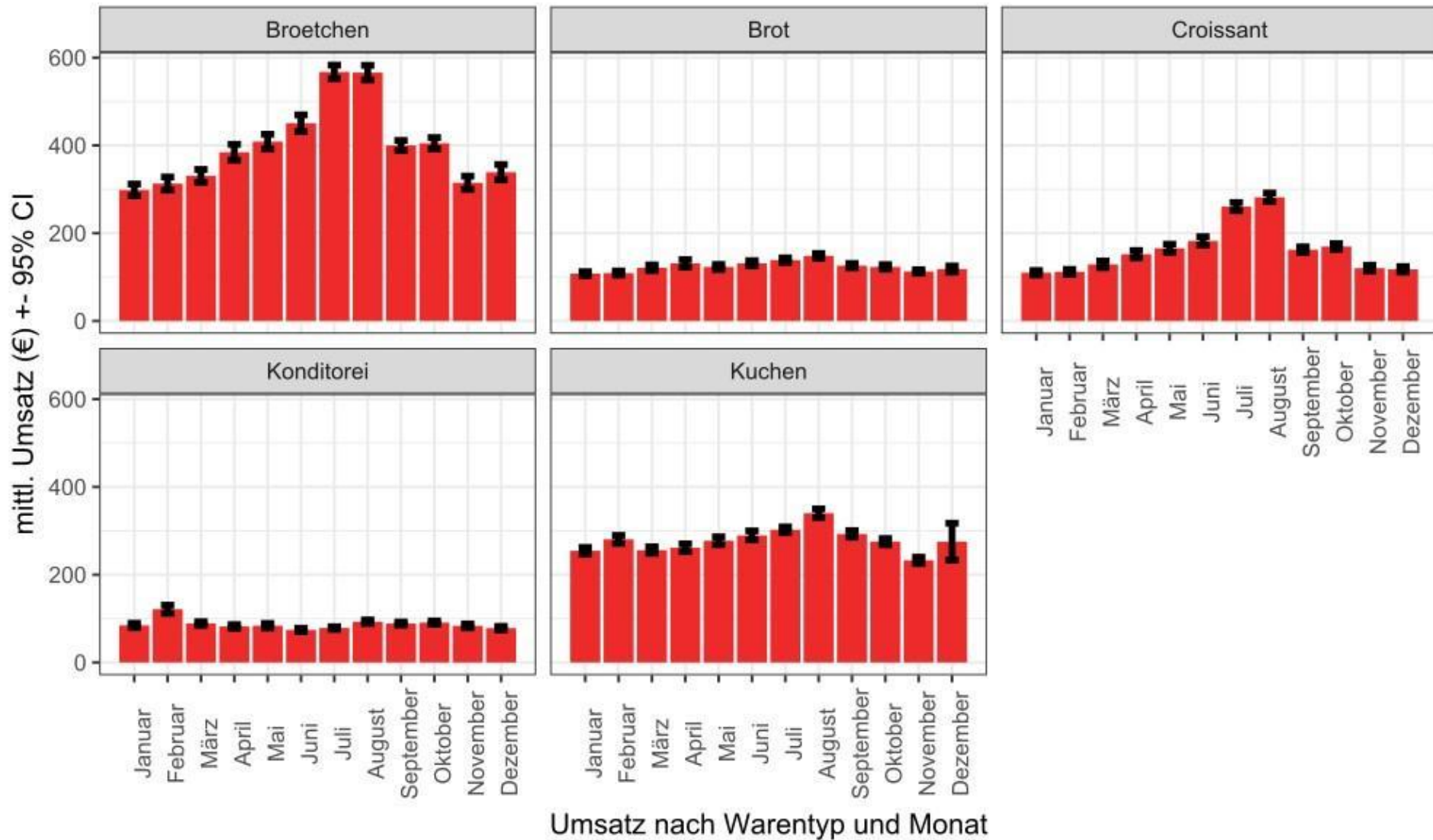
## Erstellte Variablen:

- Wochentag, Monat, Jahr
- Ferientage und Feiertage
- Tag im Monat; Monatsanfang, -mitte, -ende

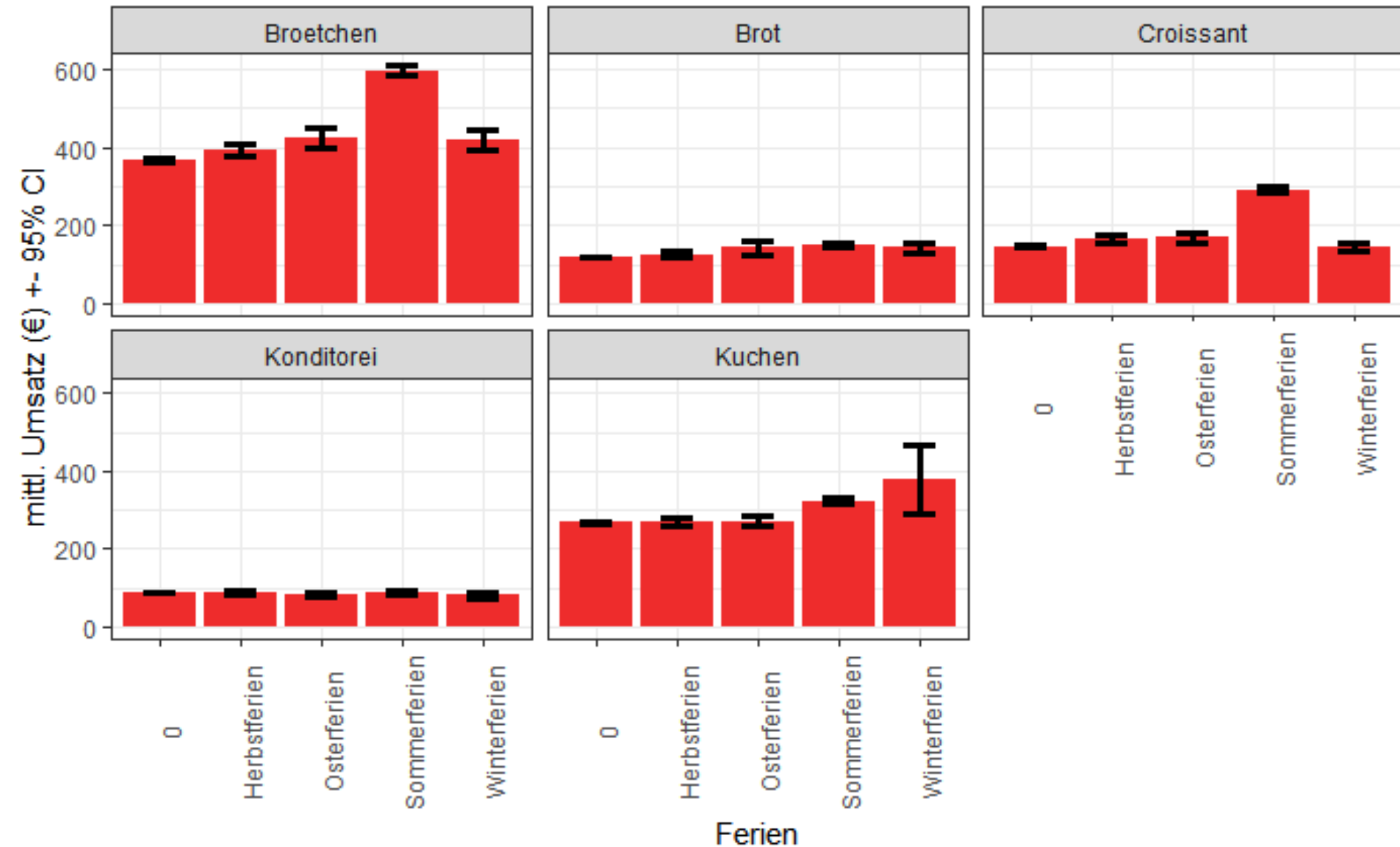
## Bearbeitungsschritte:

- Ersetzen des Code für Warentyp, Wettercode
- Unterschiedliche Zusammenfassung von Wettercode Informationen
- Löschen aller NA Werte beim Umsatz
- Entfernung von Warentyp "Saisonbrot"

# Erhöhte Umsätze während des Sommers



# Erhöhte Umsätze während bestimmter Ferien





# Schätzung durch eine Support Vector Maschine

## Gewähltes lineares Modell:

- Umsatz  $\sim$  Warentyp \* Monat + Wochentag + Jahr + Ferien + KielerWoche + Temperatur
- epsilon 0.2, cost 8, gamma 0.01282

MAPE Trainingsdaten: 16,55 %

MAPE Testdaten: 17,76 %

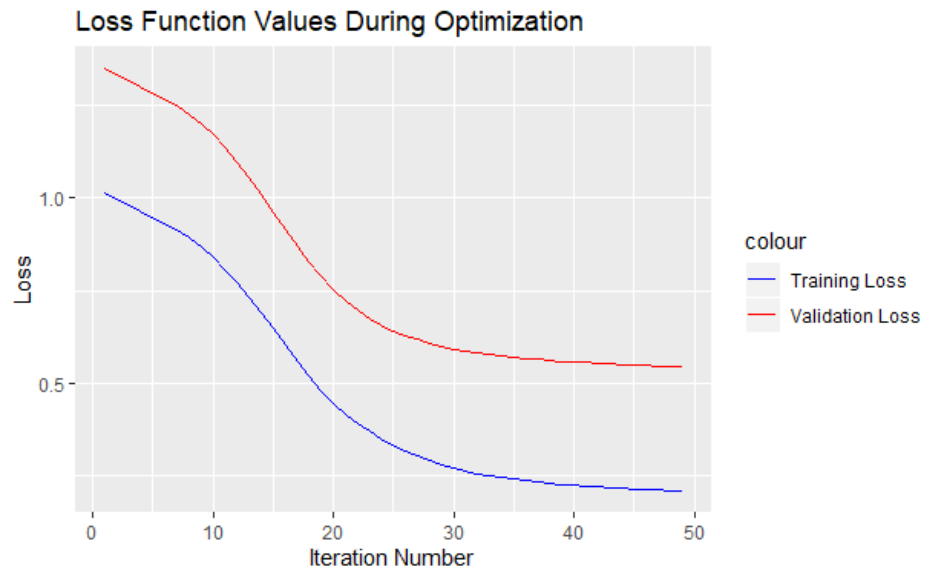
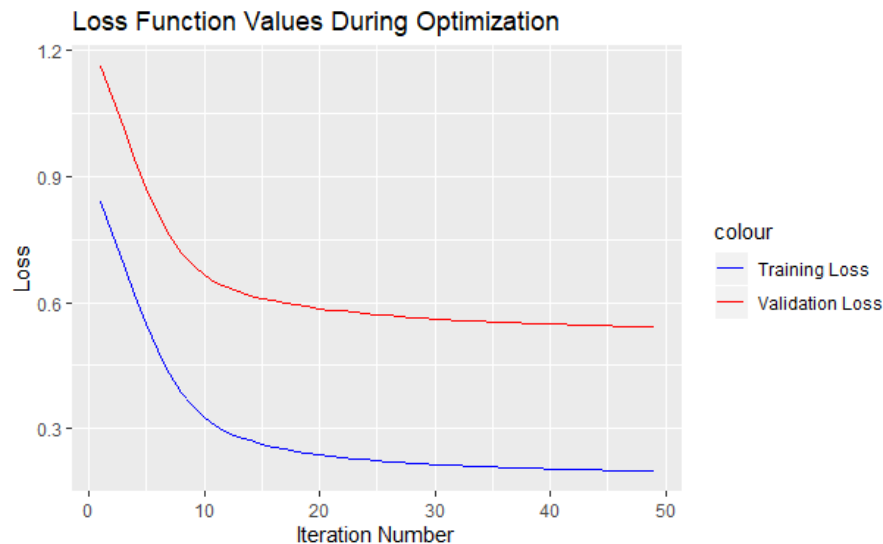
- Vergleich: MAPE Gesamtdaten fürs LM: 24,24 %

# Schätzung durch ein neuronales Netz

- Gleiche Input-Variablen wie bei SVM
- 5 Nodes als Input, 1 Hidden Layer mit 6 Nodes
- Learning Rate  $1e-4$ , Momentum= 0.9
- 50 Epochs

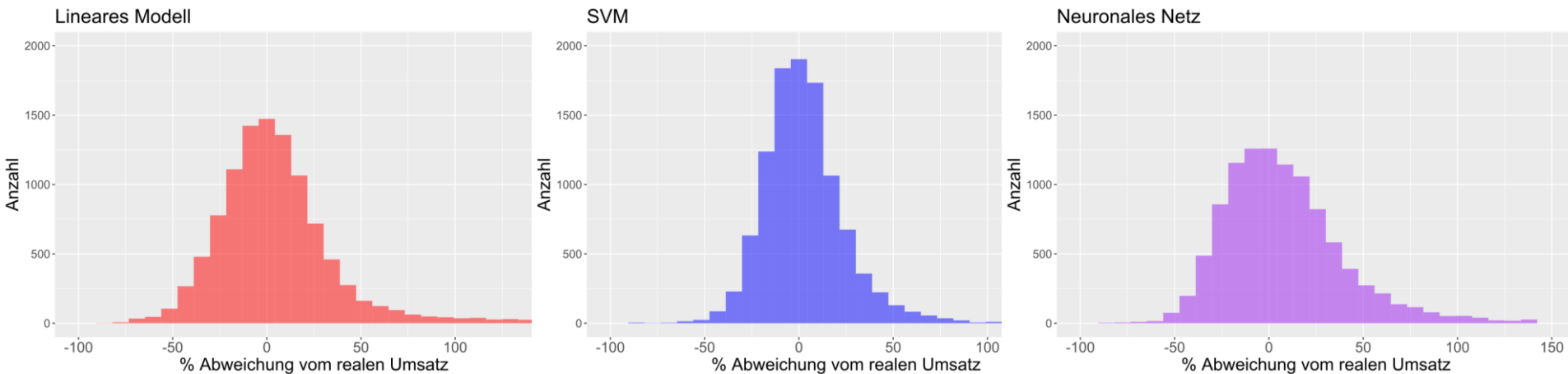
MAPE Trainingsdaten: 26,30

MAPE Testdaten: 24,87



# SVM führt zur geringsten relativen Abweichung

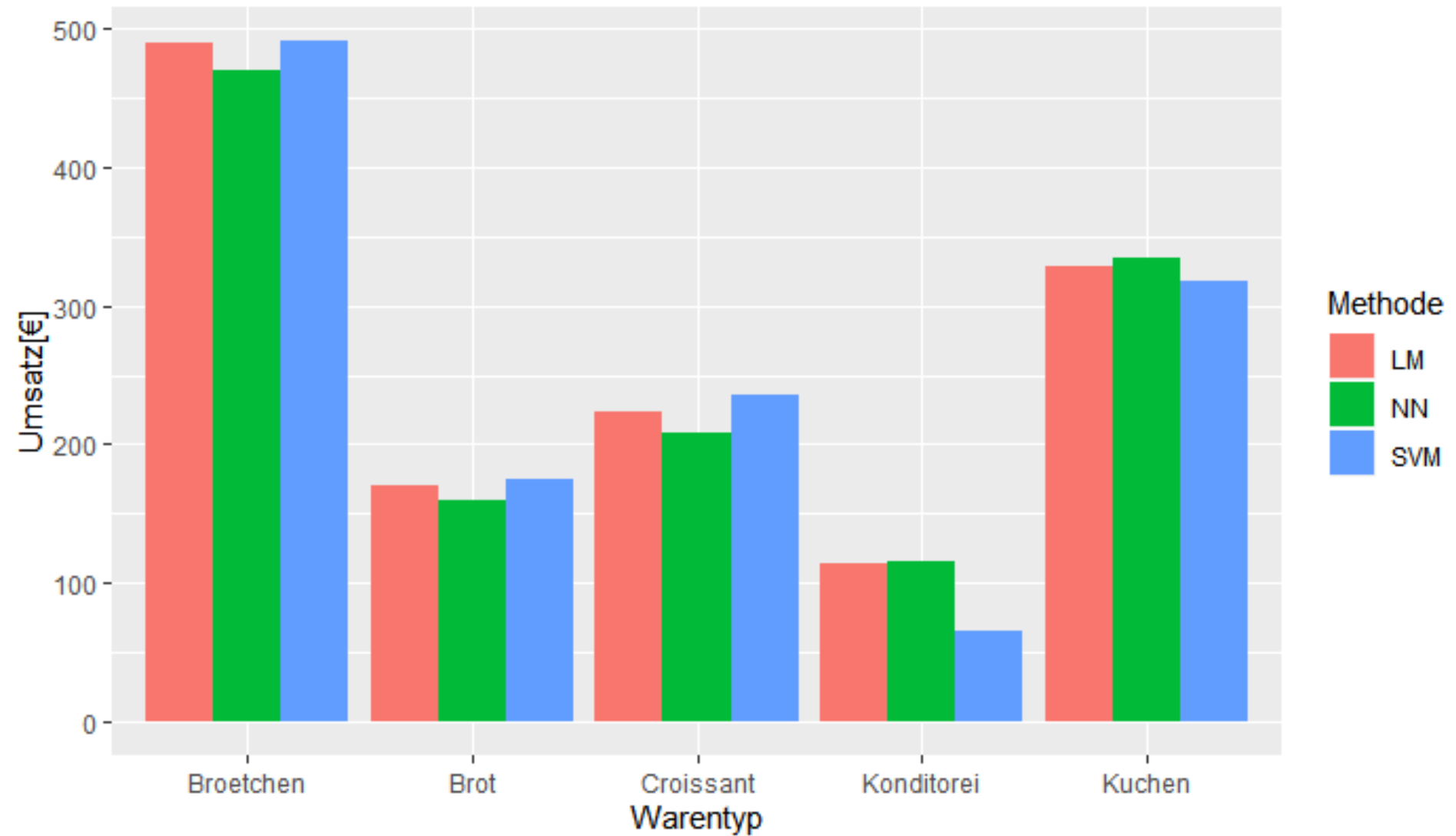
$$\% \text{ relative Abweichung} = ((\text{Voraussage} / \text{Realer Umsatz}) - 1) * 100\%$$



> 10 % rel. Abweichung (+-)

- LM: 7053 / 10441
- SVM: 6046 / 10441
- NN: 7568 / 10441

## Umsatzvorhersage für den 01.06.2019





# Vorhersagen im Vergleich

Warentyp	Umsatz LM [€]	Umsatz SVM [€]	Umsatz NN [€]
Brot	170,67	174,26	158,83
Broetchen	489,23	490,77	469,40
Croissant	223,63	236,30	208,63
Konditorei	113,96	64,99	115,64
Kuchen	328,43	318,32	334,63

# Fazit

Einordnung Güte SVM und neuronales Netz

- sehr ähnliche Umsatzergebnisse für den 1.6.
- SVM verbessert LM
- Wiederholbarkeit bei NN niedrig -> große Streuung

Ausblick

Welche weiteren Daten interessant?

- Standortabhängige Variablen
- Zielgruppe

Wie kann die Schätzung verbessert werden?

- Mikro - Zusammenhänge
- Optimierung NN: Wetterdaten mit aufnehmen?