

# Zwei-Stufen-XGBoost - Experiment-Report

Experiment-ID: v2\_h4\_thr0p5pct\_strict\_newfeat

Dieses Dokument fasst die wichtigsten Parameter, Datenquellen und Metriken eines Zwei-Stufen-XGBoost-Experiments zusammen.

Stufe 1 (Signal): neutral vs. Bewegung ('move'). Stufe 2 (Richtung): down vs. up – nur an Bewegungstagen.

## **Label-Parameter:**

- horizon\_days: 4
- up\_threshold: 0.005
- down\_threshold: -0.005
- strict\_monotonic: True

## **Datensatz & Splits:**

- dataset\_path: /Users/jeremynathan/Documents/GitHub/hs2025\_ml\_project/hs2025\_ml\_project/data/processed/datasets/eurusd\_news\_training.csv
- test\_start: 2025-01-01
- train\_frac\_within\_pretest: 0.8

**Features (FEATURE\_COLS): vollständige Liste auf der Feature-Seite weiter unten.**

# Legende & Begriffe

## Zielvariablen:

- label: 3-Klassen-Ziel auf Basis des 4-Tage-Lookaheads (neutral / up / down).
- signal: 0 = neutral, 1 = Bewegung (up oder down).
- direction: 0 = down, 1 = up; nur definiert, wenn signal == 1.

## Wichtige Metriken:

- precision: Anteil der vorhergesagten positiven Fälle, die wirklich positiv sind.
- recall: Anteil der tatsächlichen positiven Fälle, die erkannt wurden.
- f1: harmonischer Mittelwert aus precision und recall (Balance beider Größen).
- support: Anzahl der Beobachtungen in der jeweiligen Klasse.

## Feature-Abkürzungen (Auswahl):

- article\_count: Anzahl News-Artikel pro Tag.
- avg\_polarity / avg\_neg / avg\_neu / avg\_pos: durchschnittliche Sentiment-Werte.
- pos\_share / neg\_share: Anteil positiver bzw. negativer Sentiment-Komponente.
- intraday\_range\_pct: (High - Low) / Close – relative Tages-Spanne (Volatilität).
- upper\_shadow / lower\_shadow: obere/untere Dochte der Kerzen (High/Low vs. Körper).
- month / quarter: Kalendermonat und Quartal.

## **Modell-Parameter (XGBoost)**

### **Signal-Modell (Stufe 1):**

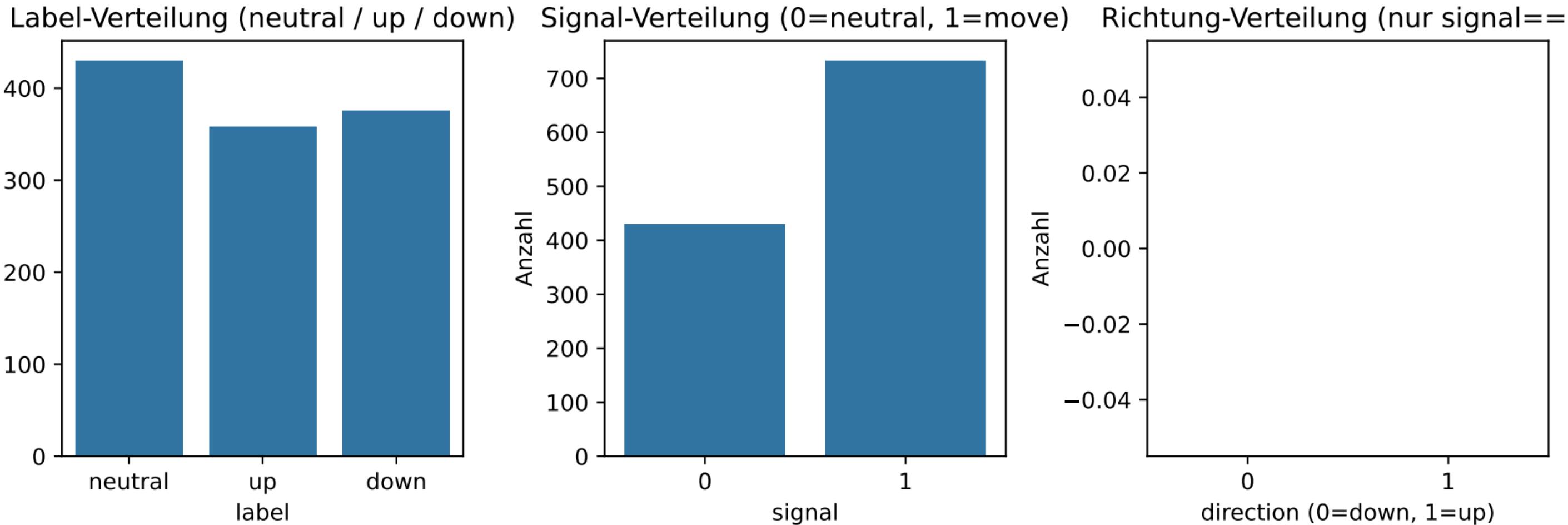
- objective: binary:logistic
- max\_depth: 3
- learning\_rate: 0.05
- n\_estimators: None
- subsample: 0.9
- colsample\_bytree: 0.9
- scale\_pos\_weight: 7.689655172413793

### **Richtungs-Modell (Stufe 2):**

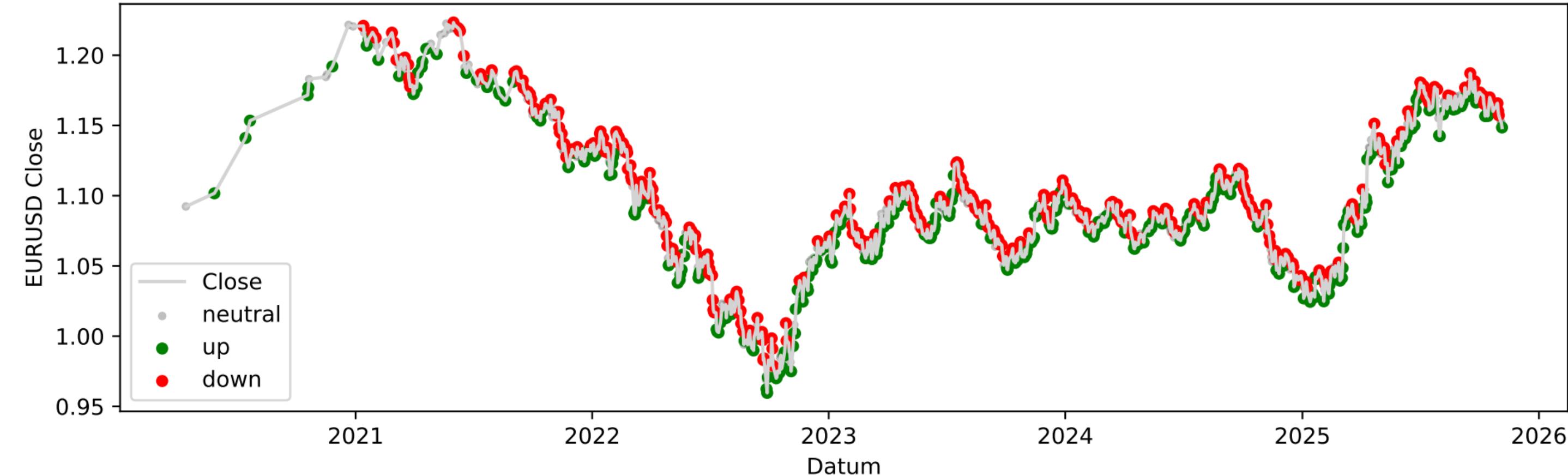
- objective: binary:logistic
- max\_depth: 3
- learning\_rate: 0.05
- n\_estimators: None
- subsample: 0.9
- colsample\_bytree: 0.9
- scale\_pos\_weight: 1.0

## Verwendete Features (FEATURE\_COLS)

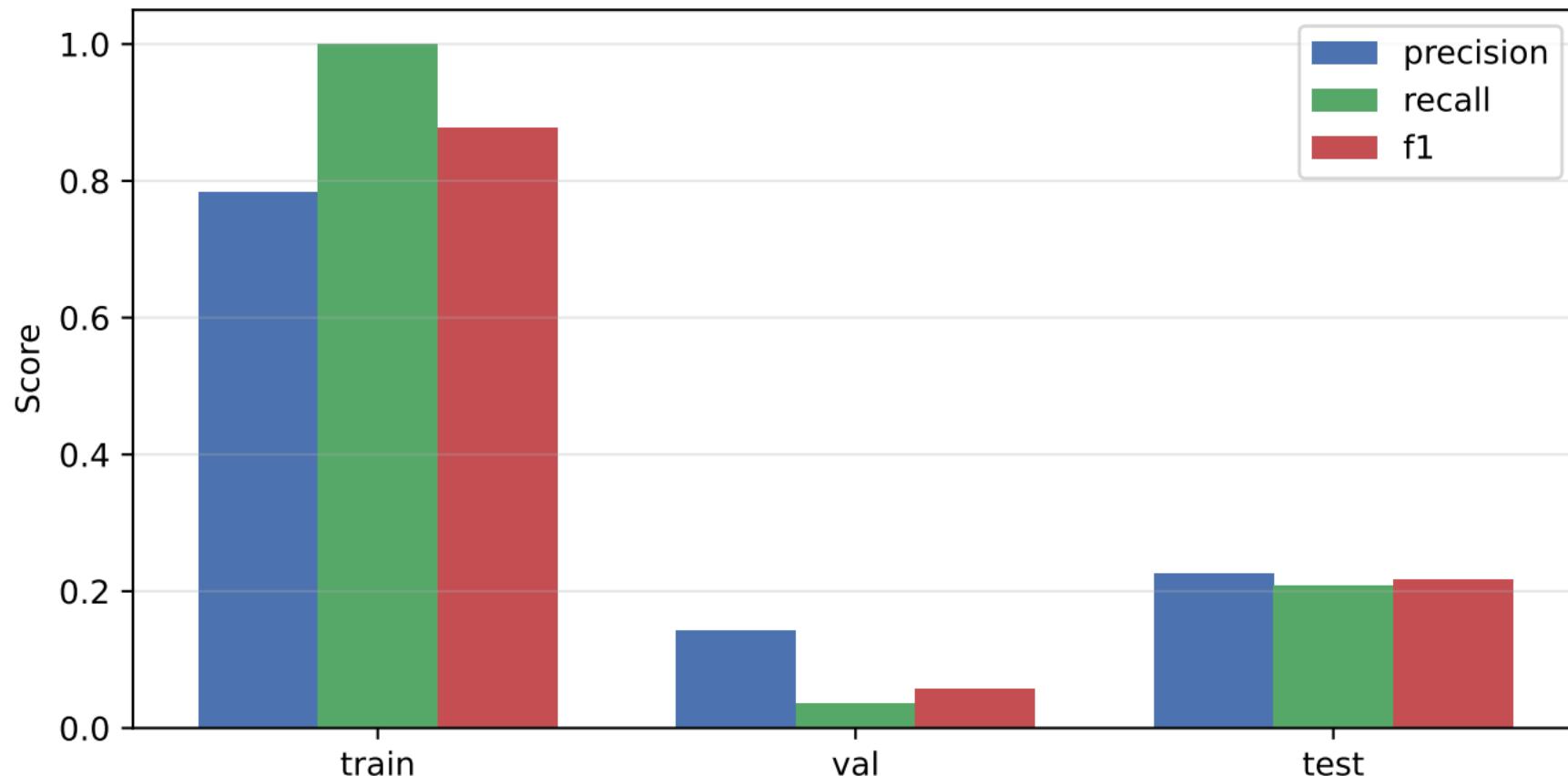
#	feature_name
0	article_count
1	avg_polarity
2	avg_neg
3	avg_neu
4	avg_pos
5	pos_share
6	neg_share
7	intraday_range_pct
8	upper_shadow
9	lower_shadow
10	price_close_ret_1d
11	price_close_ret_5d
12	price_range_pct_5d_std
13	price_body_pct_5d_mean
14	news_article_count_3d_sum
15	news_article_count_7d_sum
16	news_pos_share_5d_mean
17	news_neg_share_5d_mean
18	news_article_count_lag1
19	news_pos_share_lag1
20	news_neg_share_lag1
21	month
22	quarter
23	cal_dow
24	cal_day_of_month
25	cal_is_monday
26	cal_is_friday
27	cal_is_month_start
28	cal_is_month_end
29	hol_is_us_federal_holiday
30	hol_is_day_before_us_federal_holiday
31	hol_is_day_after_us_federal_holiday



# EURUSD-Zeitreihe mit hervorgehobenen up/down-Tagen (ab 2020)



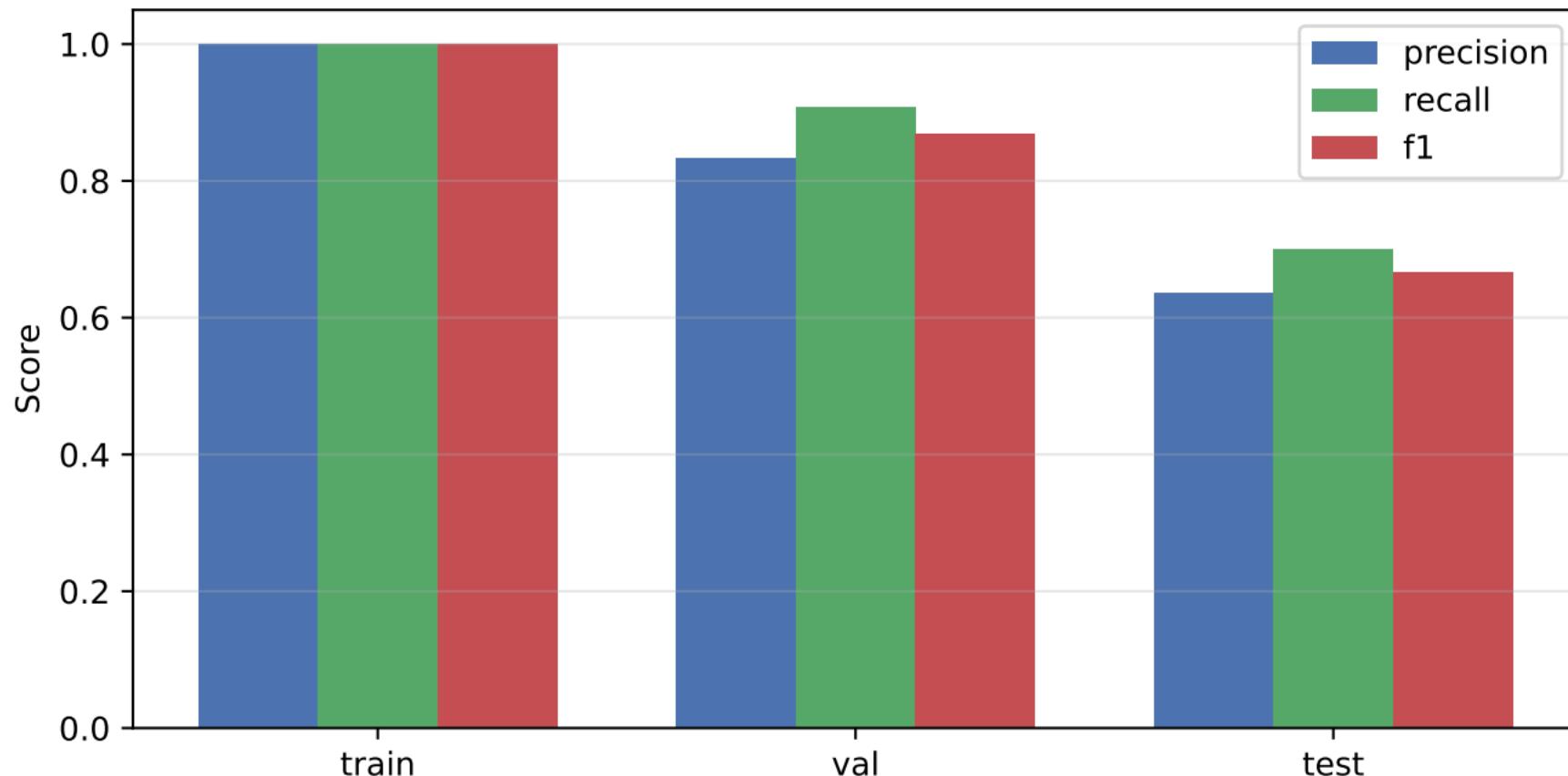
## Signal-Modell - Kennzahlen für Klasse 'move' (train/val/test)



## **Signal-Modell - Tabelle (Klasse 'move')**

split	precision	recall	f1	support
train	0.784	1.000	0.879	87.000
val	0.143	0.036	0.057	28.000
test	0.227	0.208	0.217	24.000

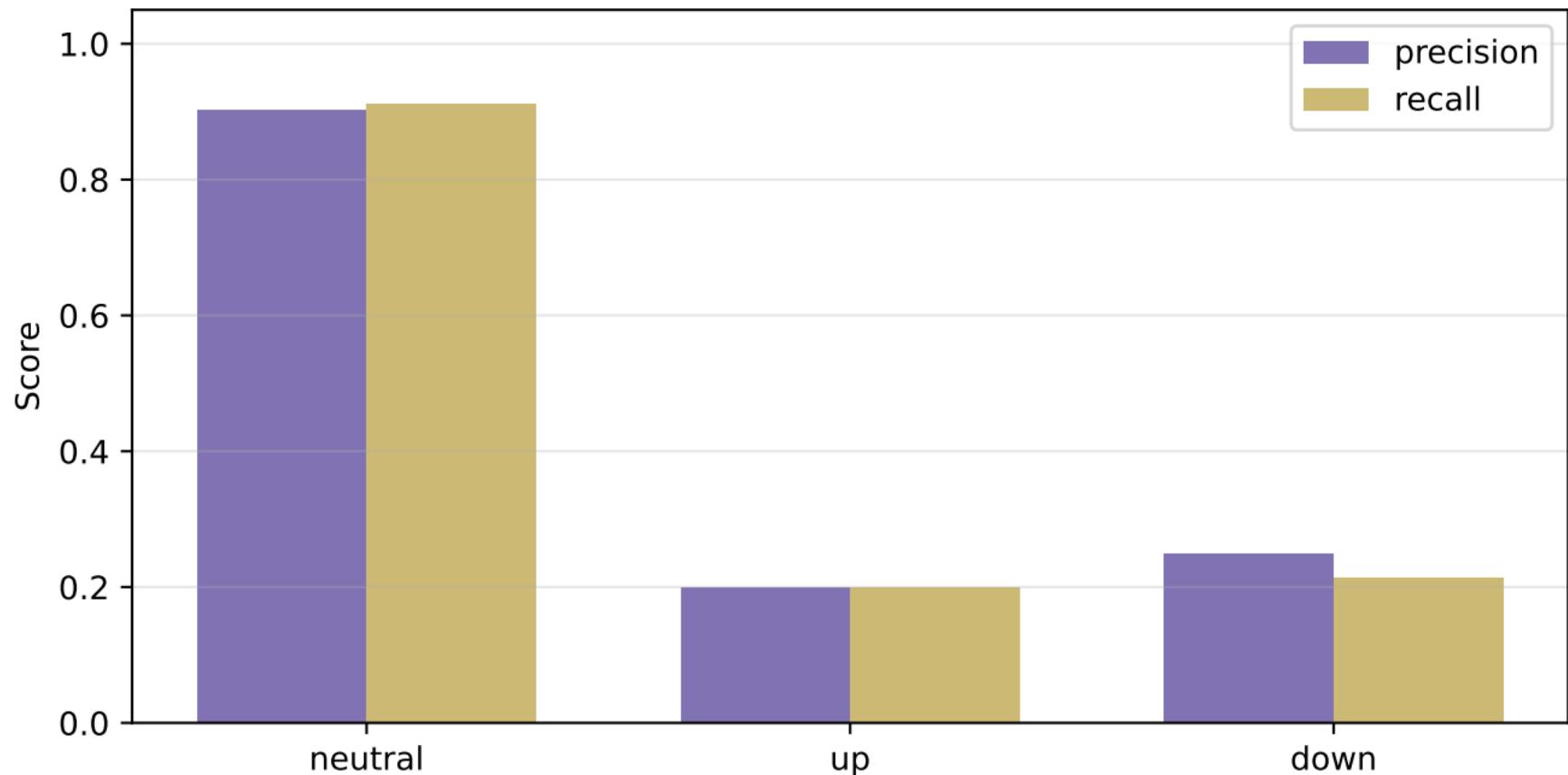
## Richtungs-Modell – Kennzahlen für Klasse 'up' (train/val/test)



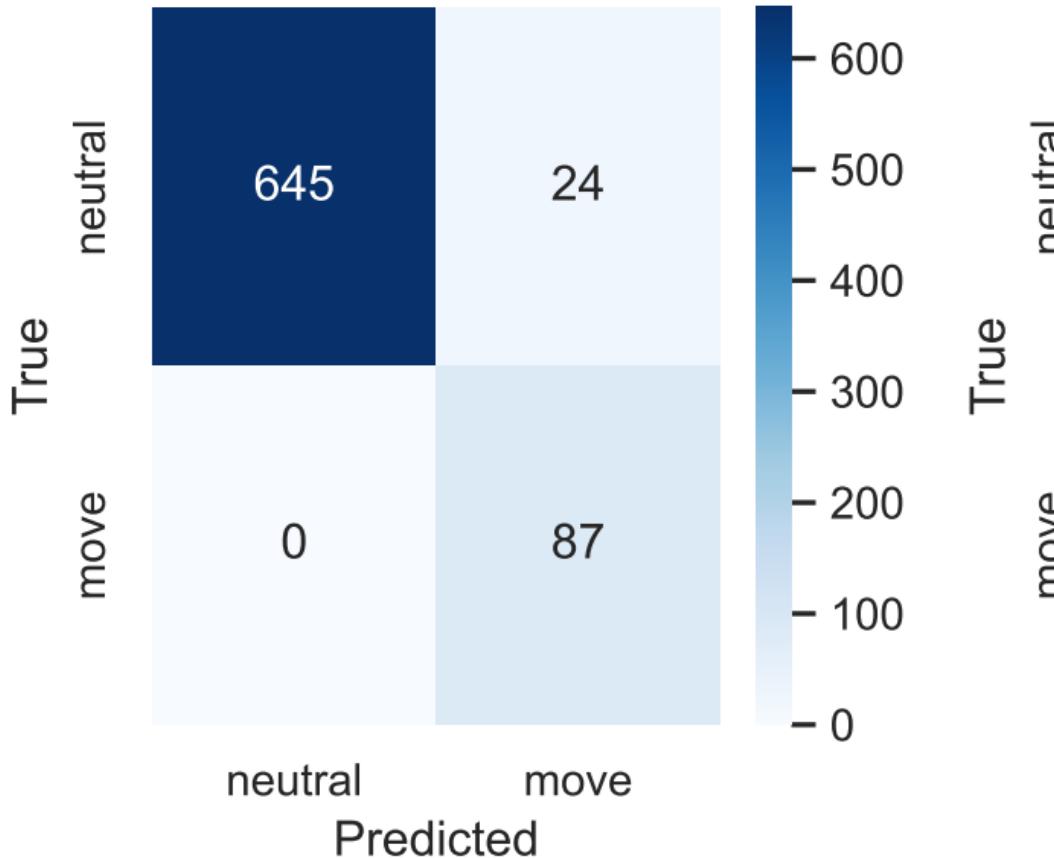
## Richtungs-Modell - Tabelle (Klasse 'up')

split	precision	recall	f1	support
train	1.000	1.000	1.000	38.000
val	0.833	0.909	0.870	11.000
test	0.636	0.700	0.667	10.000

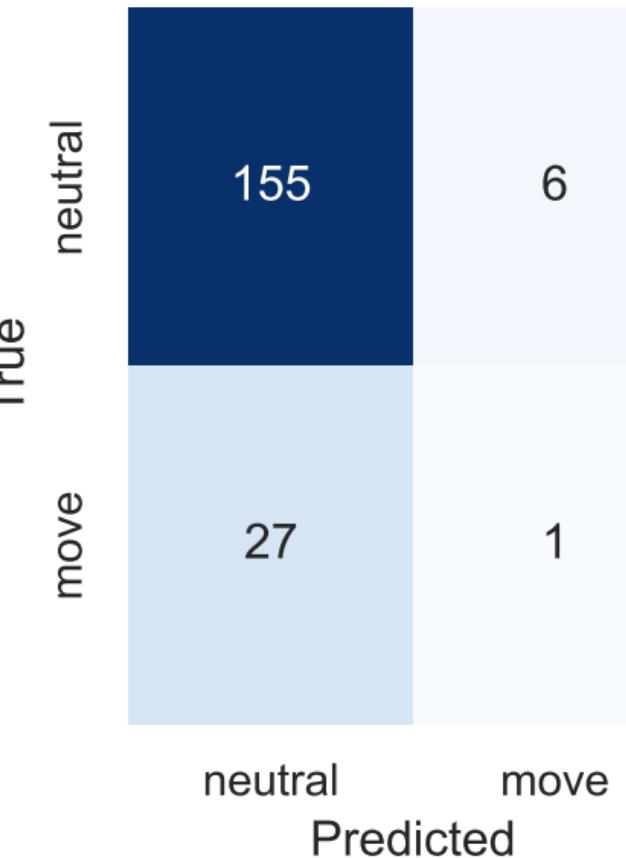
## Kombinierte Test-Auswertung - neutral / up / down



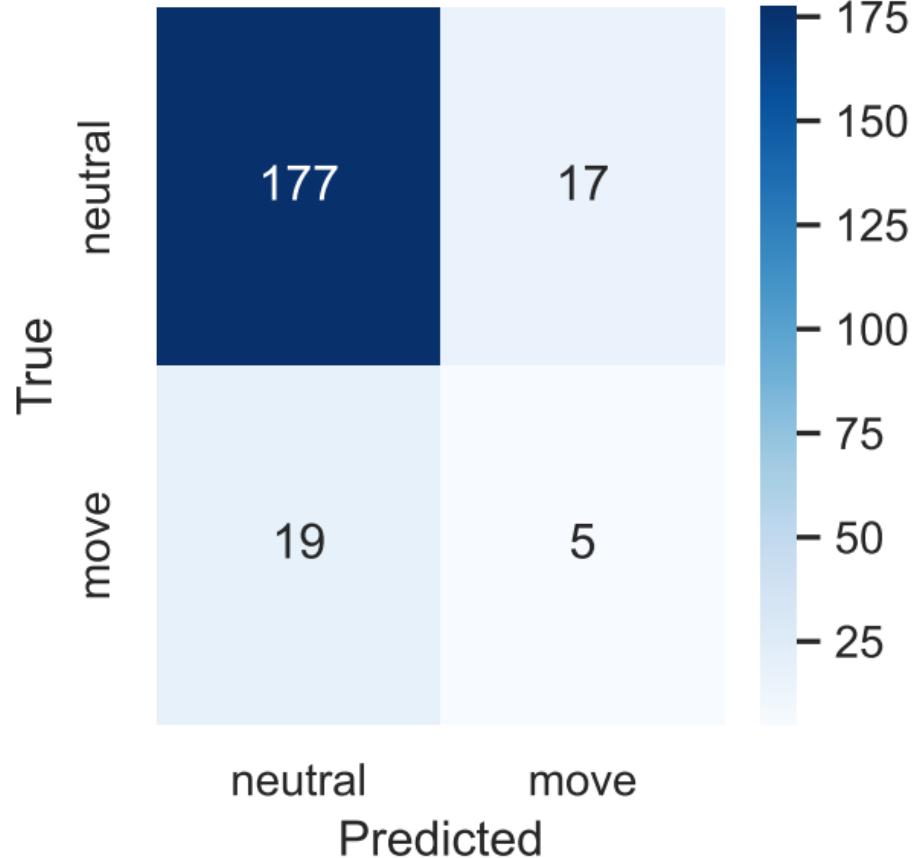
Signal – train



Signal – val



Signal – test



Richtung – train

True  
down

49

0

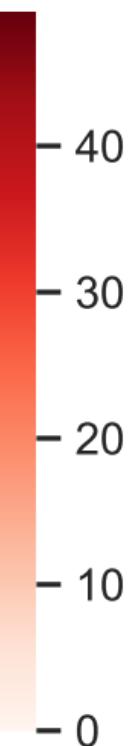
up

0

38

down

Predicted



Richtung – val

True  
down

15

2

up

1

10

down

Predicted



Richtung – test

True  
down

10

4

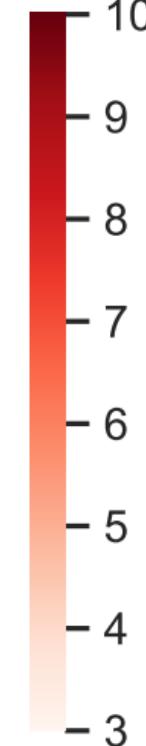
up

3

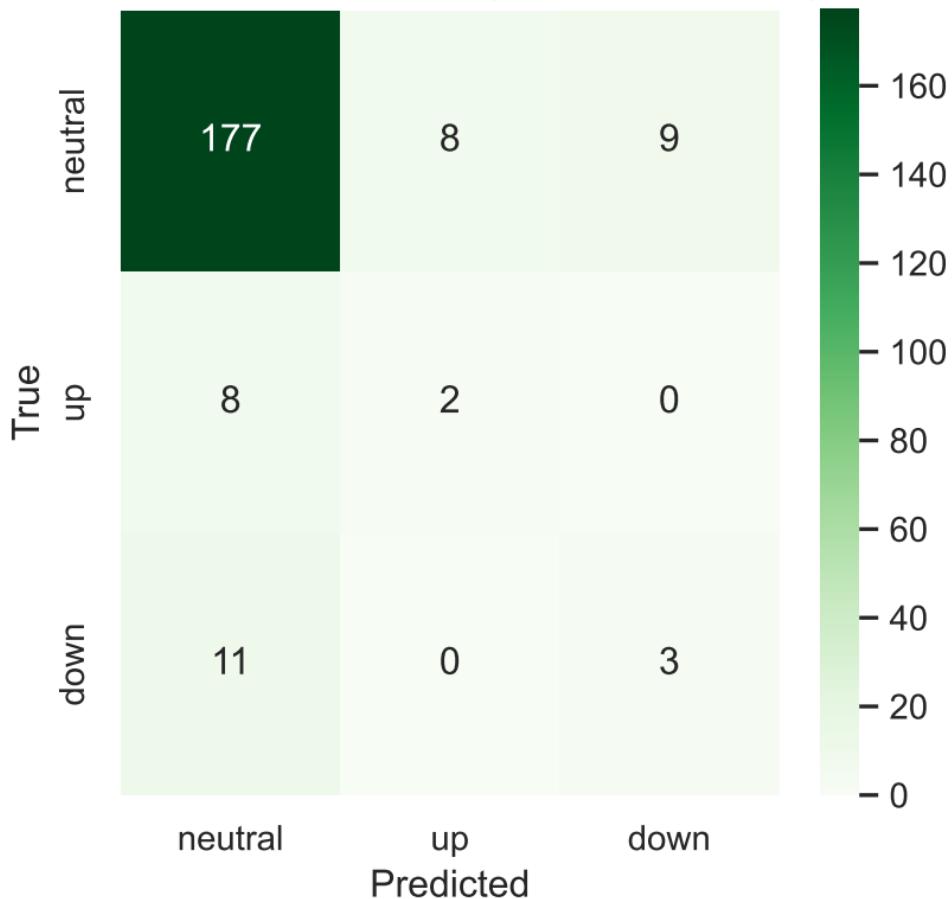
7

down

Predicted



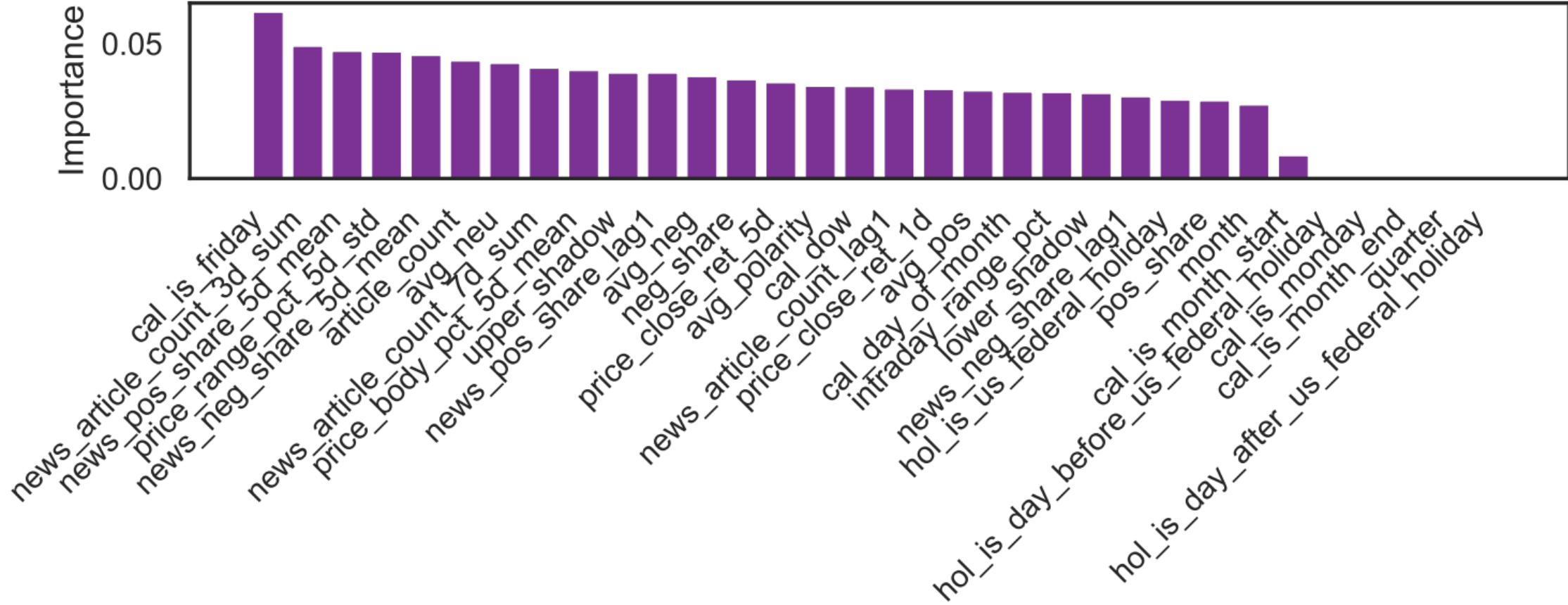
# Confusion Matrix – Test (neutral / up / down)



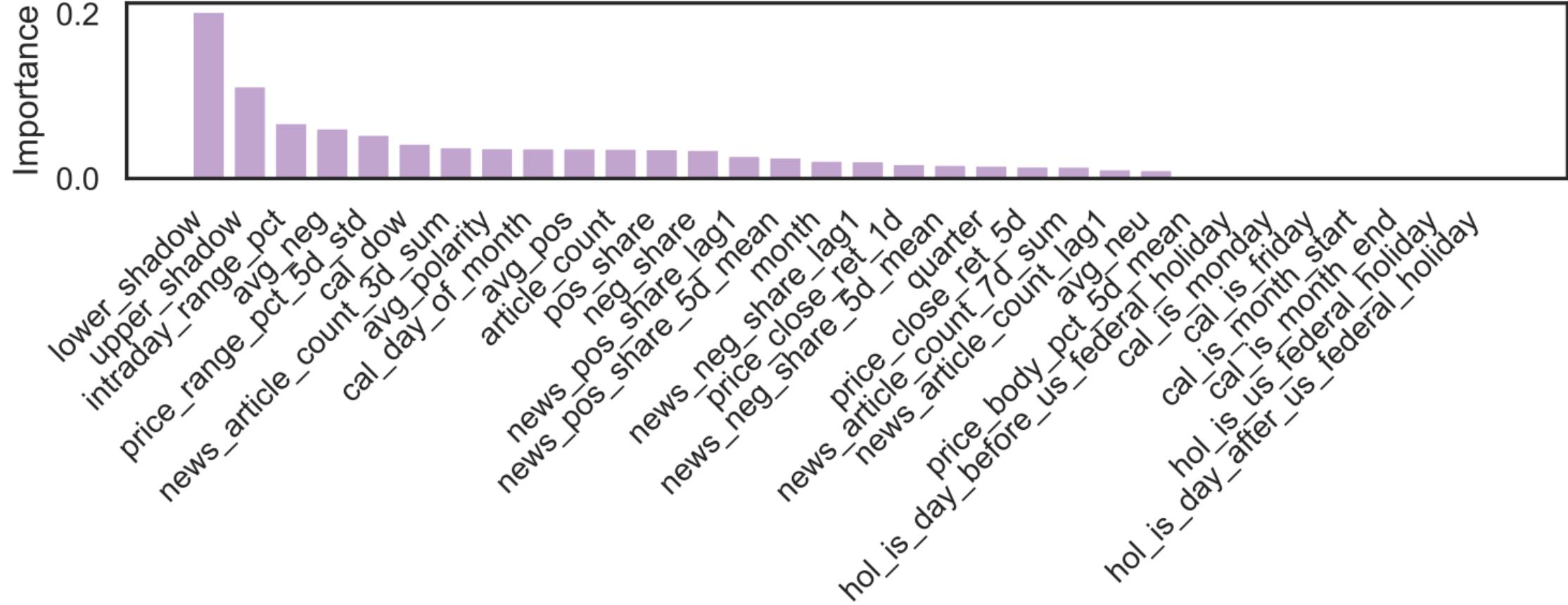
## Konfusionsmatrizen – Zählwerte (TN/FP/FN/TP)

modell	split	TN	FP	FN	TP
signal (neutral vs. mov	train	645	24	0	87
signal (neutral vs. mov	val	155	6	27	1
signal (neutral vs. mov	test	177	17	19	5
direction (down vs. up	train	49	0	0	38
direction (down vs. up	val	15	2	1	10
direction (down vs. up	test	10	4	3	7

## Feature Importance – Signal-Modell



## Feature Importance – Richtungs-Modell



Confusion Matrix – Test (Richtung: down vs. up)

