Predicting Conversions from Print Newspaper Subscription Sales

Analysis for a German publisher (Frankfurter Allgemeine Zeitung)

METIS - Classification

Fabian Paul, 13/06/2022

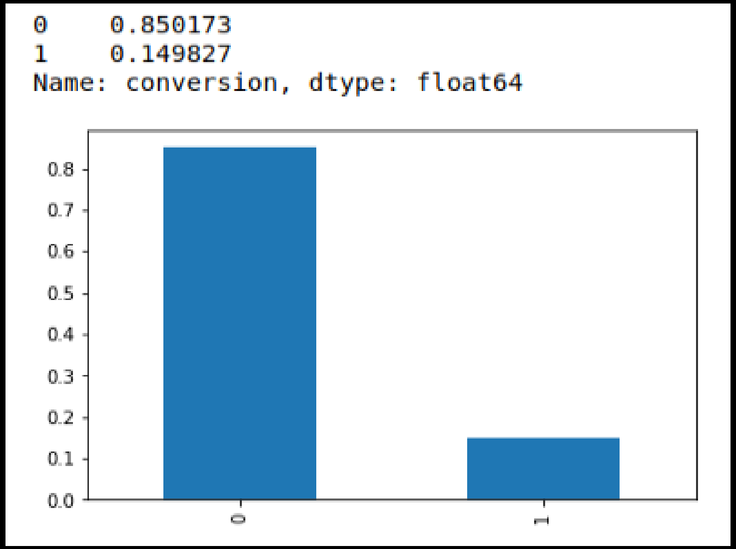
Introduction

Business Problem

- Print subscriptions most important revenue source
- Converted Subs = Regular Paid Subscriptions with ≥ 90 days duration
- Prediction model crucial for:
  - … identifying features that drive conversions
  - … target right customers with “care actions”

Objective

- ➢ Which order, customer and product features are significantly correlated to conversions?
- ➢ Is it possible to predict conversions instantaneously when an order is received?



Methodology

Order Data

13,015 Transactions Q3/2021

- *Orders*
- *Products*
- *Customers*

Feature Engineering

- *Convert subscription data (SAP) into usable format*
- *Create bins of broader categories*
- *Generate target variable for each subscription*
- *Generate features on past customer relationships*
- *Convert categories to dummies*

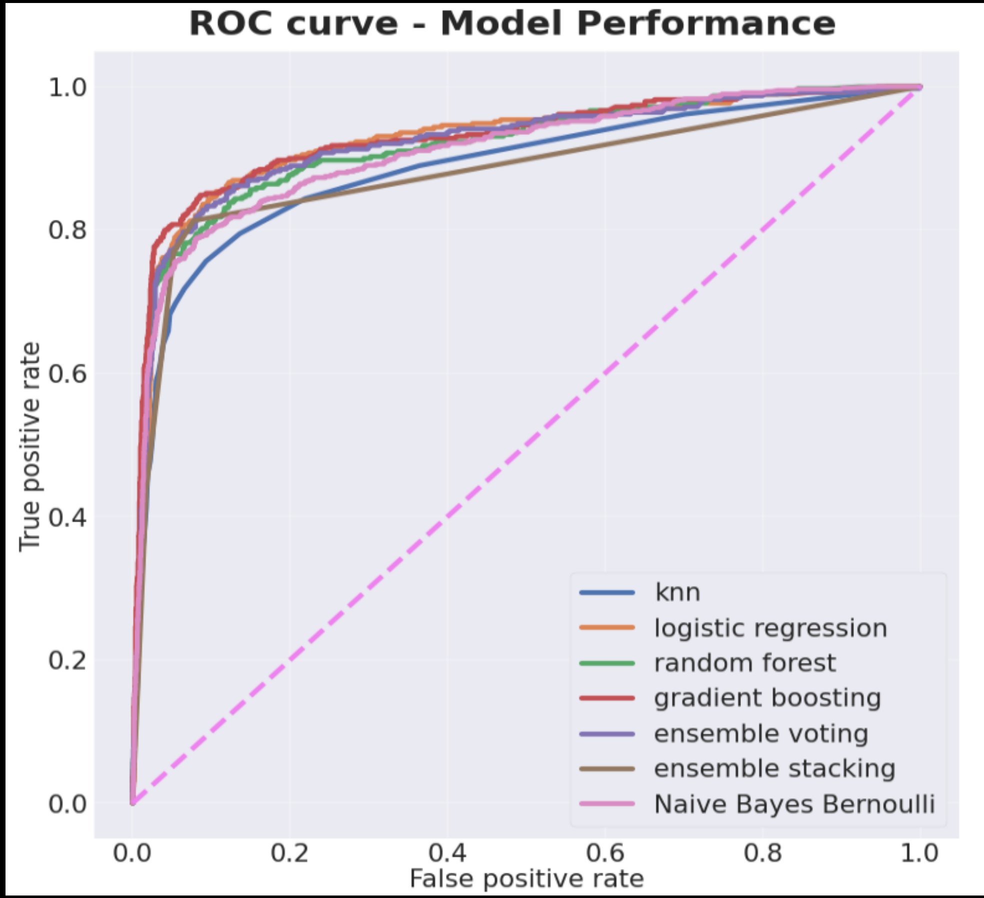
liefende	12269
gultig_bis	11886
erfdat_kun	2354
aart	13015
posart	13015
bezgrd	13015
preisgruppe	5207
bezgrd_lfzt	13015
bezgrd_option	13015
faktura_period	13015
werbeart	13015
abgangs_typ	13015
lieferart	13015
zahlweg	13015
sachpramie	13015
rg	5210
we	13015
we_anrede	13015
we_optin_email	13015
we_optin_tel	13015
we_optin_brief	13015
amount	13015
we_title_cat	13015
age_cat	5930
cust_type	13015
marketing_channel	13015
rg_we	13015
state	13015
conversion	13015
offer_id	13015
offer_rebate	9479
days_since_last_order	7265
days_since_first_order	13015
no_orders_twohalfyears	13015
no_orders_oneyr	13015
no_orders_halfyr	13015
active_subs_atorder	13015
active_digsubs_atorder	13015
active_sundayprint_atorder	13015
days_since_last_enddate	13015
days_since_last_churndate	13015
diff_start_order	13015
order_day	13015
type	13015

Source: Frankfurter Allgemeine Zeitung GmbH

Results

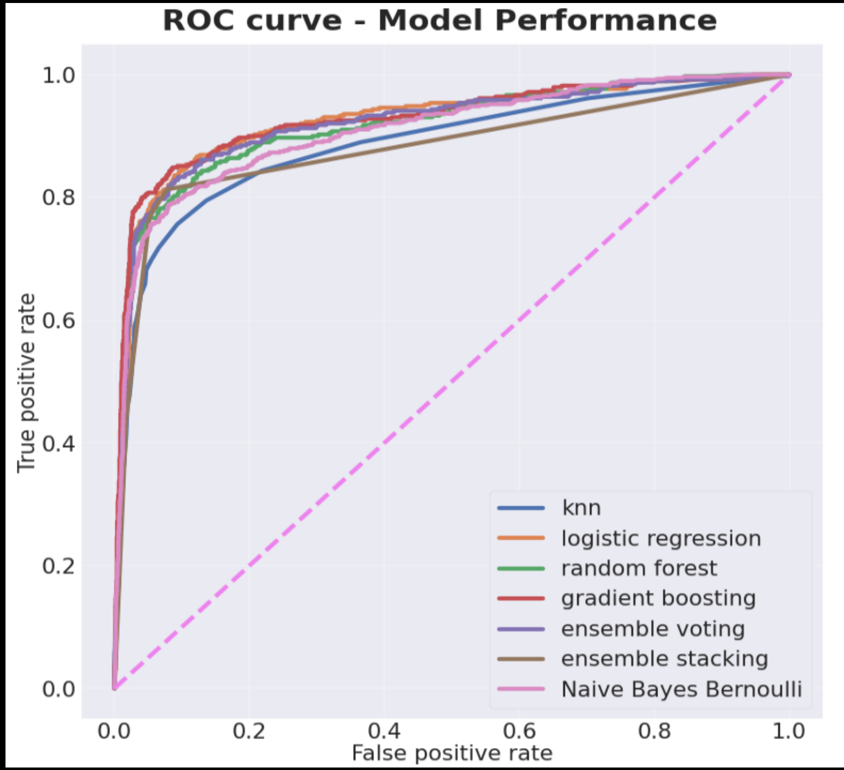
Model Evaluation

```
Precision Scores:
Gradient Boosting =  0.8385269121813032
Logistic Regression =  0.6509240246406571
Ensemble Voting =  0.7730870712401056
Random Forest =  0.6810933940774487
Naive Bayes Bernoulli =  0.738831615120275
KNN =  0.8104265402843602
Ensemble Stacked =  0.7228915662650602
```



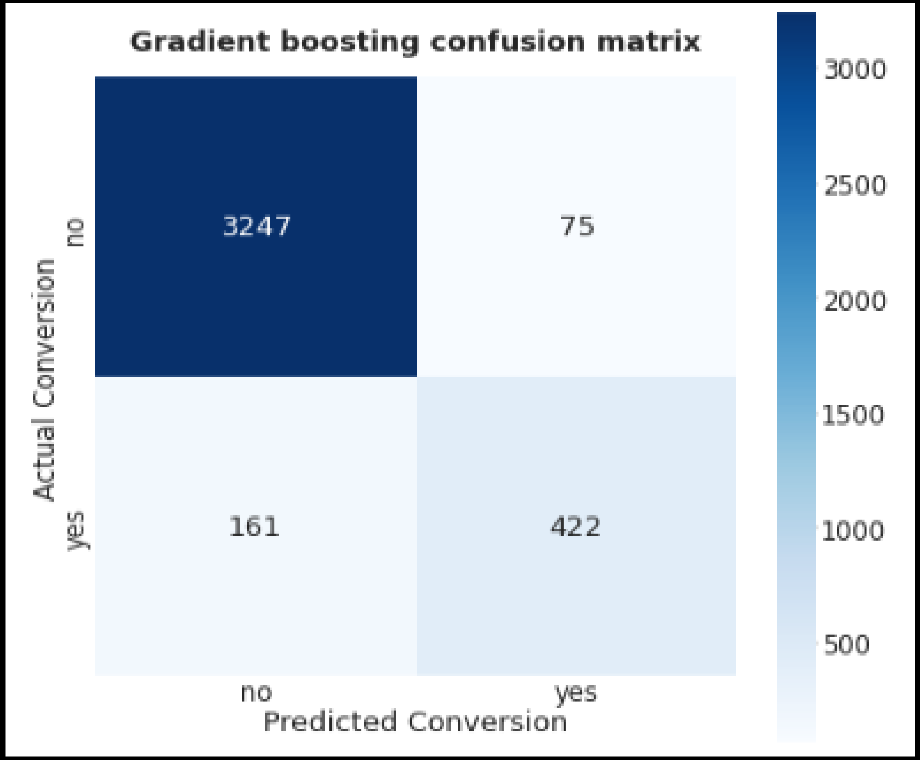
Results

	Precision	F1 Score	Recall	ROC AUC	Logloss
KNN	0.81	0.57	0.44	0.90	0.43
Logistic Regression	0.65	0.72	0.81	0.93	0.30
Random Forest	0.68	0.72	0.77	0.92	0.29
Gradient Boosting	**0.84**	0.80	0.76	0.94	0.19
Naive Bayes	0.74	0.74	0.74	0.91	0.33
Ensemble Stacking	0.72	0.75	0.77	0.88	0.23
Ensemble Voting	0.77	0.76	0.75	0.93	0.25

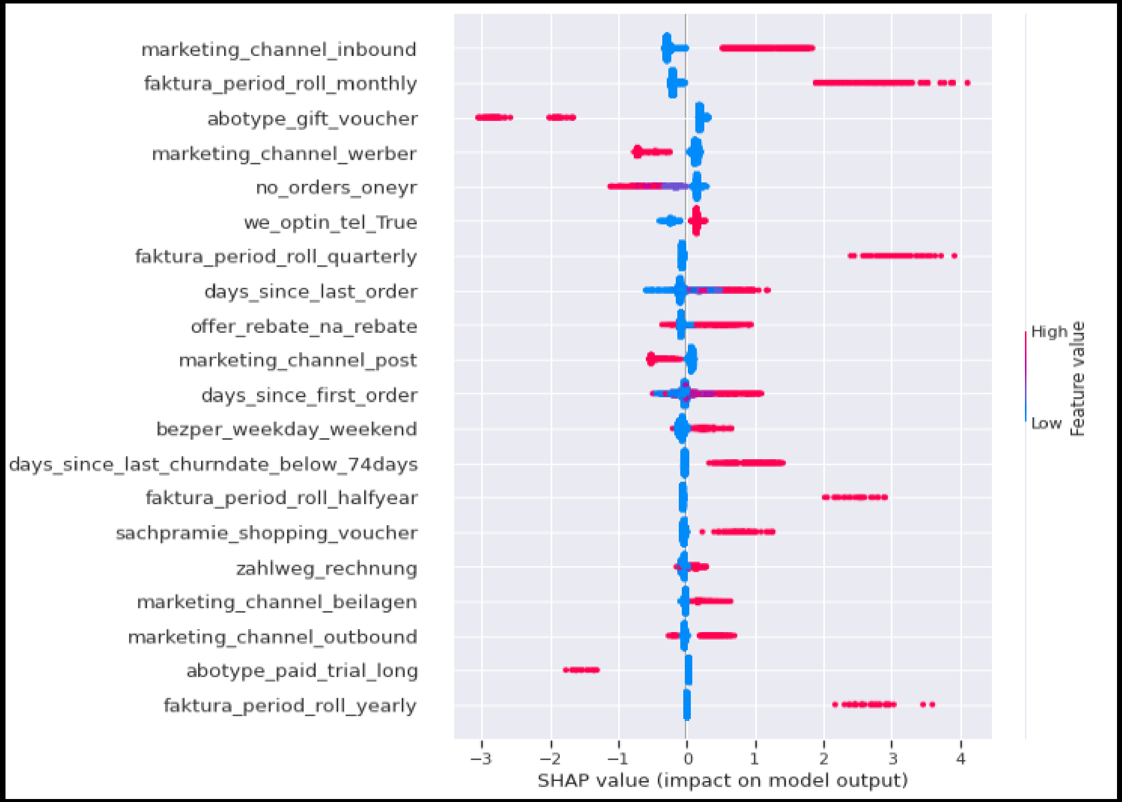


Results

Confusion Matrix



Feature Importance



Conclusions

Takeaways

- Order Features:
  - Marketing channel, payment period, shopping voucher
- Customer features:
  - No. of orders in last year, days since first & last order, optin-approval
- It is possible to <u>predict conversions</u> right after a subscription sale with high precision

Limitations

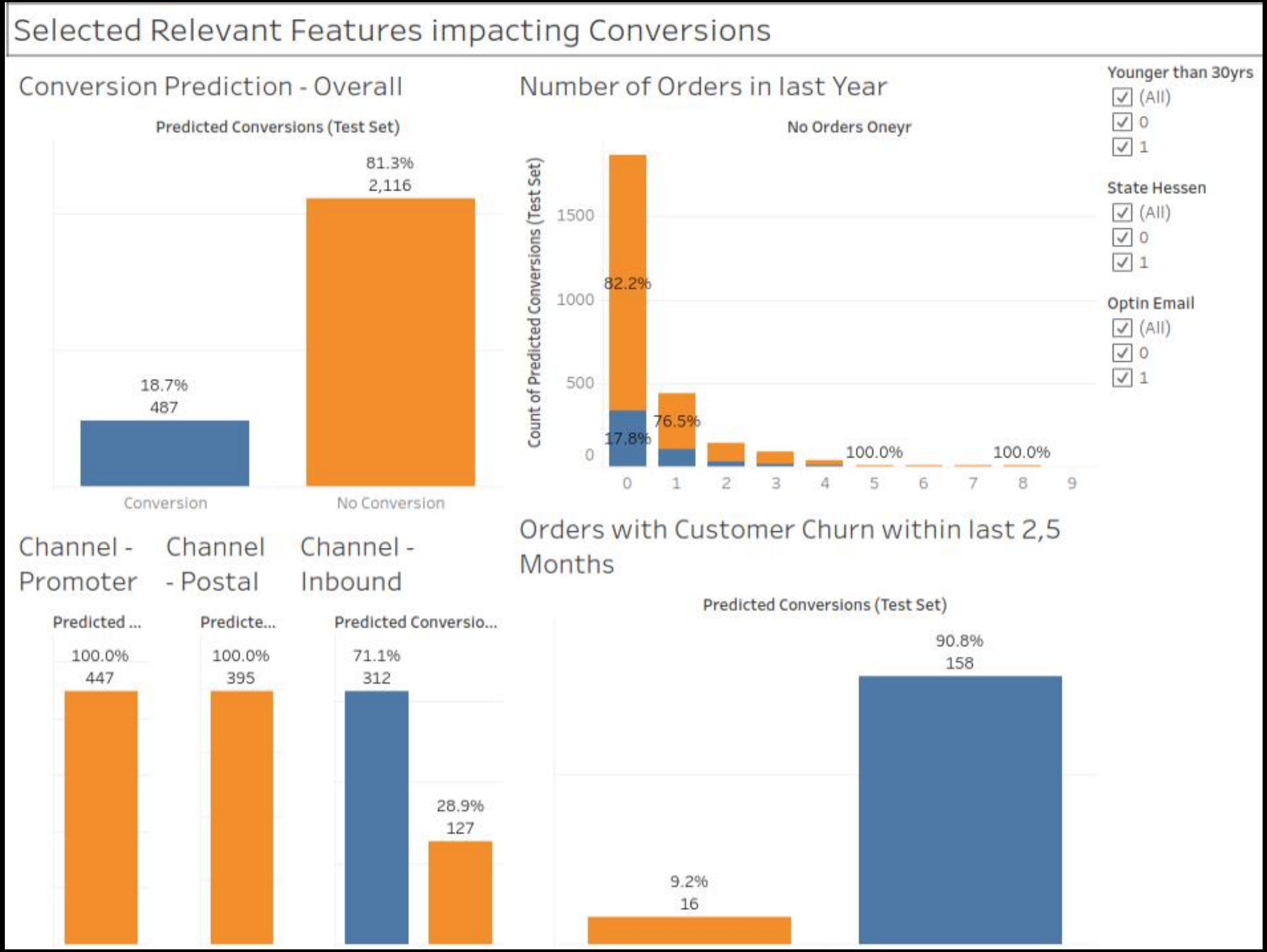
- Only Weekday Print Subscriptions
- Prediction only at point of order (not afterwards)
- No information on digital product usage
- No information on customer service data
- Socio-demographic data not available for all records

Future Work

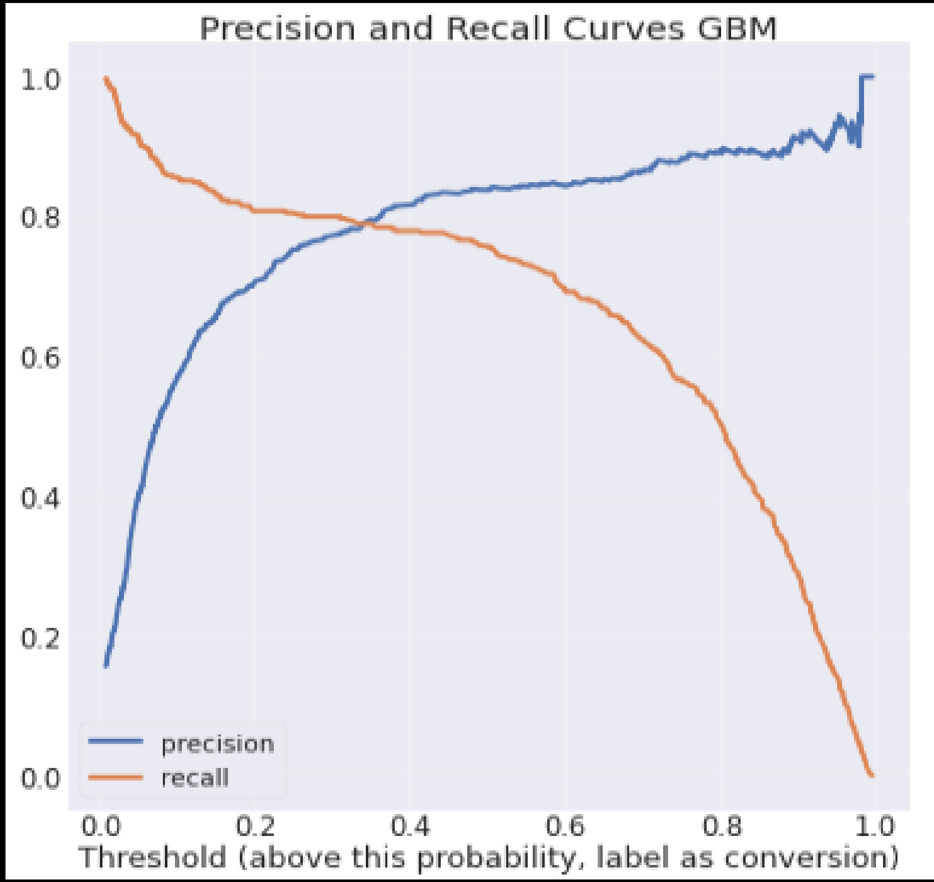
1. → Train model on more products besides weekday print
2. → Include digital product usage and customer service data
3. → Fit model each day to account for different subscription durations at time of training
4. → Interactive Web-Interface
5. → Exclusion of not meaningful features, automated model selection & code efficiency

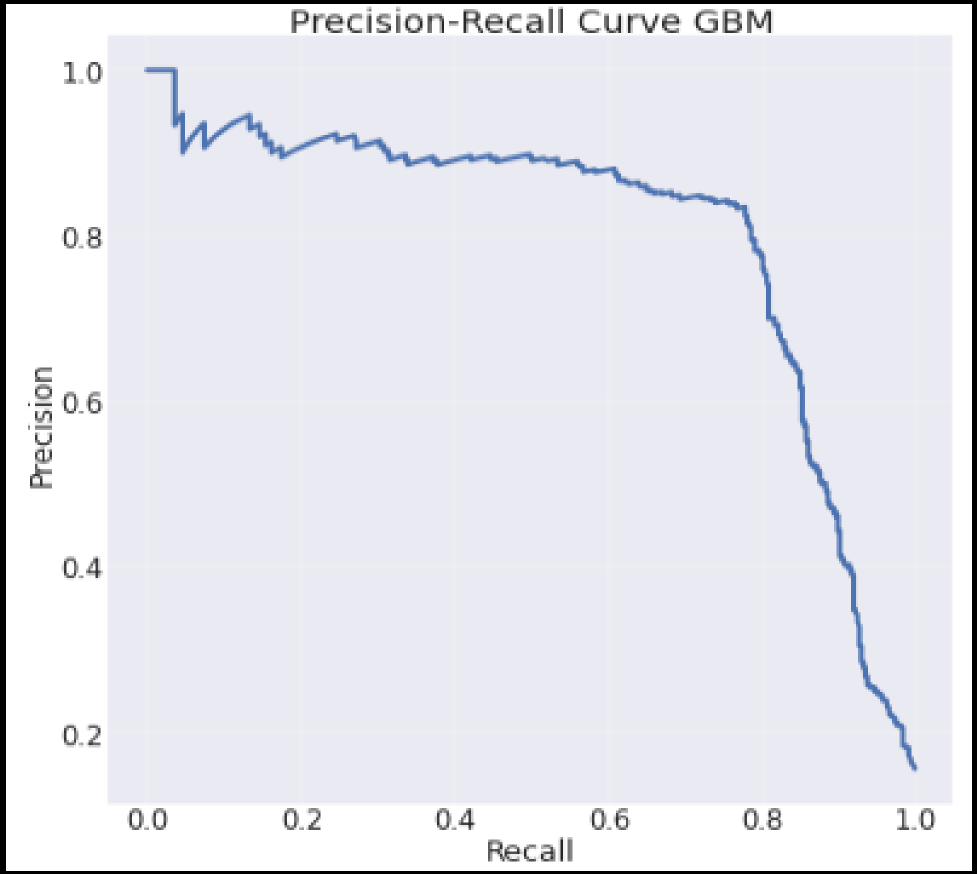
Thank you for your attention!

Appendix I: Selected Features (Tableau)



Appendix II: Precision and Recall Curves Gradient Boosting





Appendix III: Feature Coefficients Logistic Regression

	Feature	Coefficients
0	faktura_period_roll_monthly	2.588469
1	faktura_period_roll_quarterly	2.492067
2	faktura_period_roll_yearly	2.108892
3	faktura_period_roll_halfyear	2.023117
4	marketing_channel_inbound	1.699640
...	...	...
76	no_orders_oneyr	-0.601616
77	marketing_channel_post	-0.677757
78	marketing_channel_werber	-0.948955
79	abotype_paid_trial_long	-1.279328
80	abotype_gift_voucher	-3.131272

81 rows × 2 columns

Appendix IV: Correlation Heatmap