

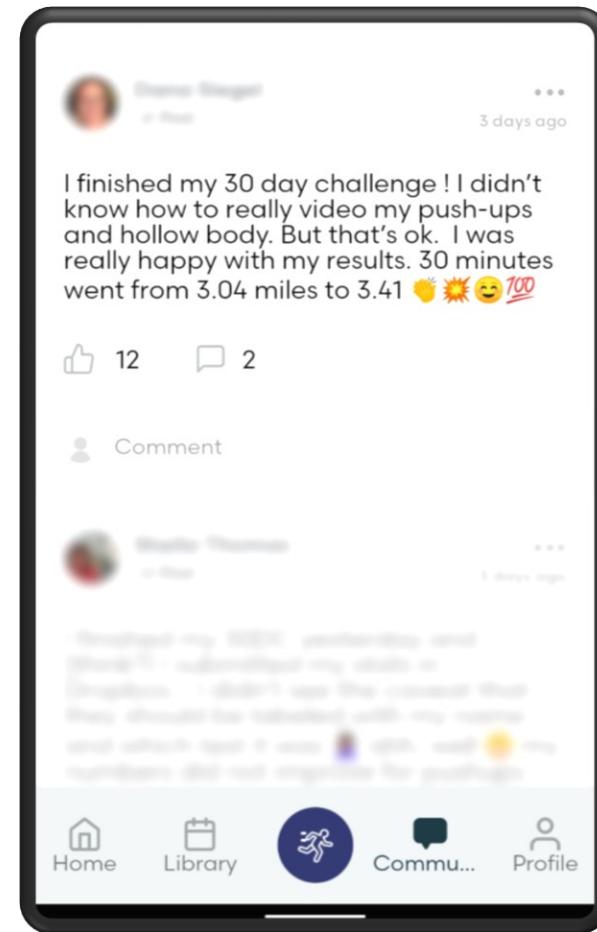
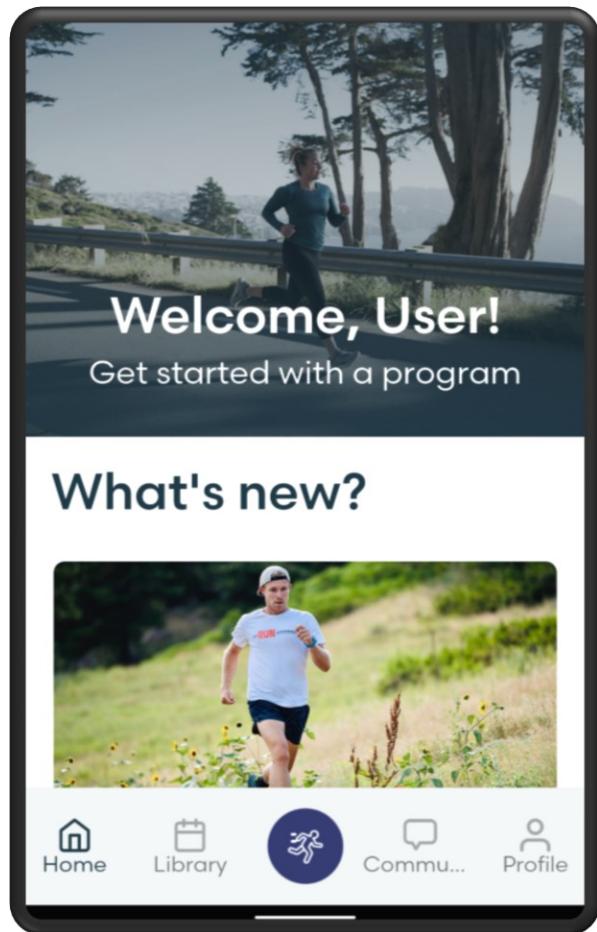
Beyond Words

predict user churn with text (meta)data

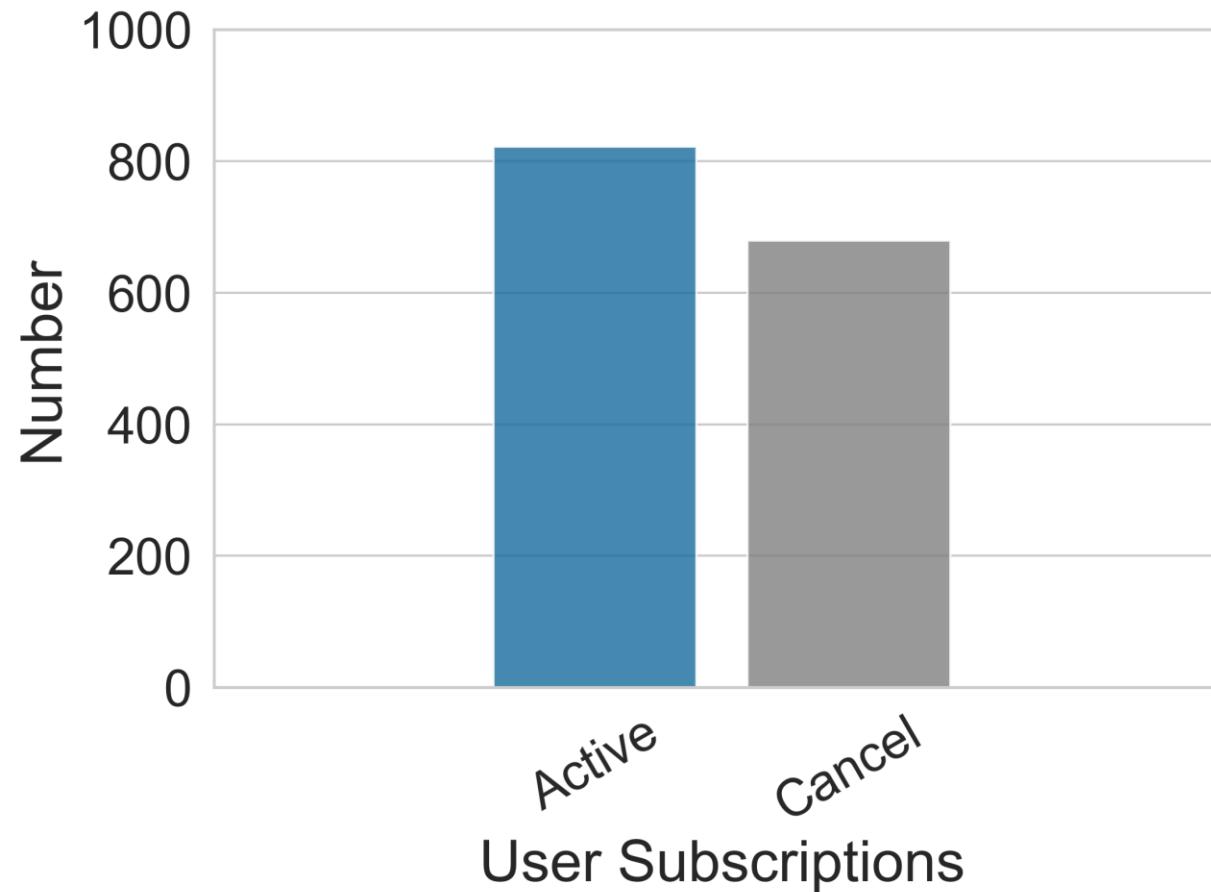
Eric Zhang



Text data from user in-app communication



45% churn, big impact to revenue



Use text data to predict user churn

Meta

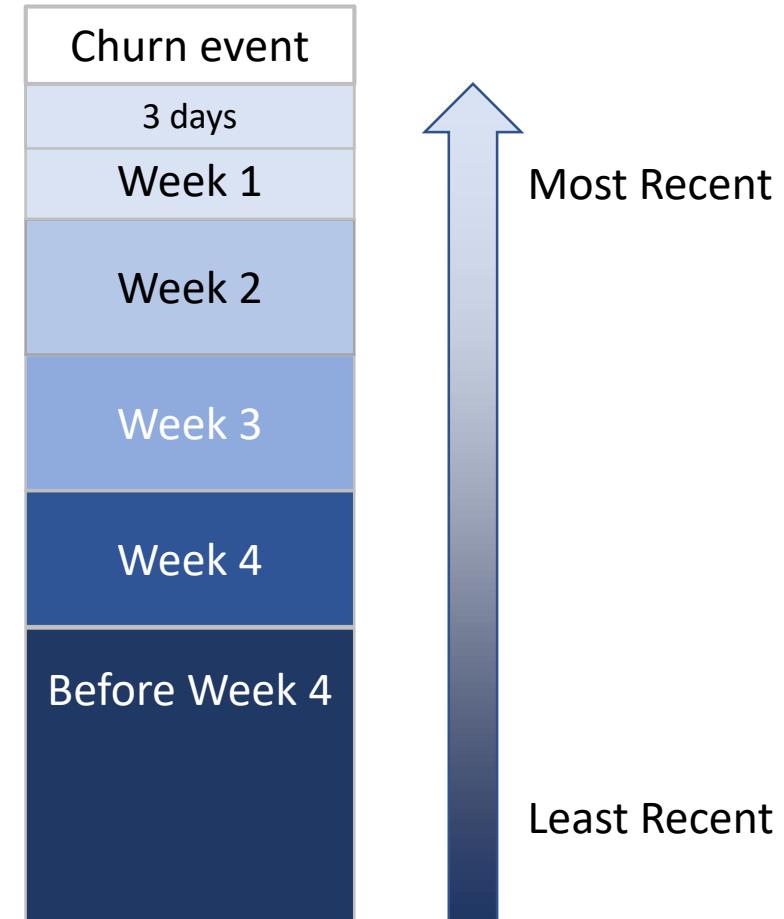
- Number of character/ text
- Likes received
- Timestamp

Sentiment

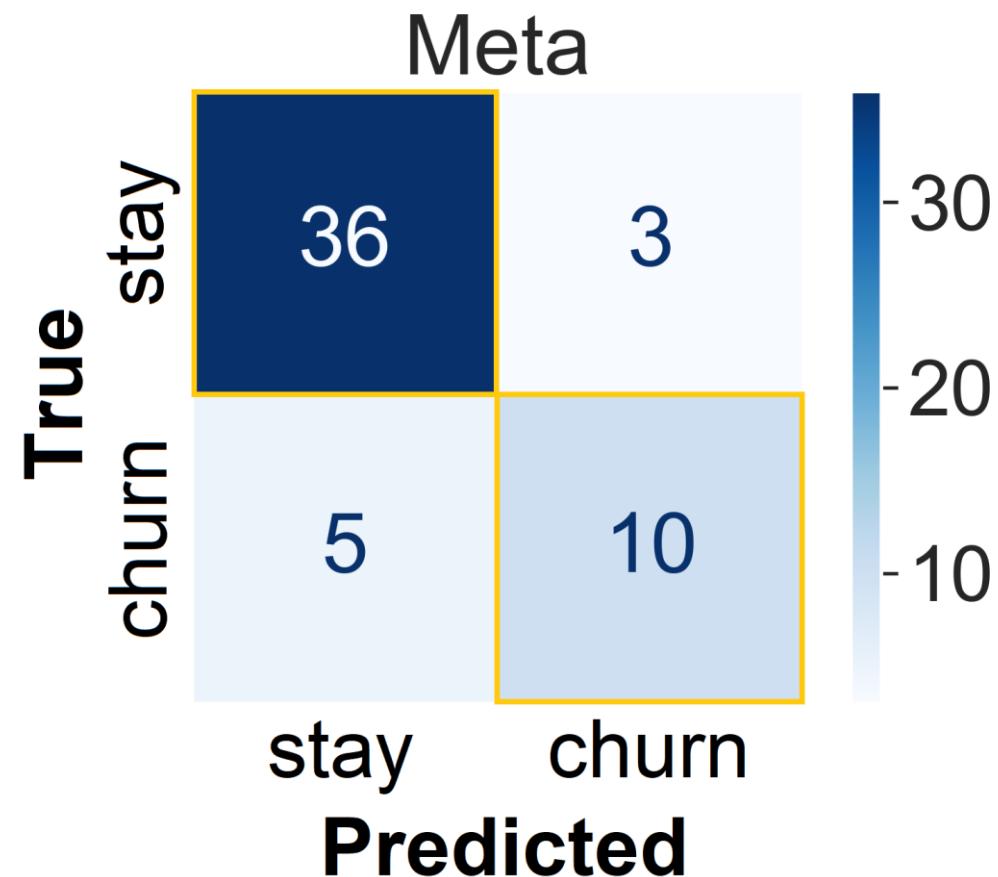
- Happy, frustrated

Text meta (characters, texts, likes)

- Before decision (churn or not)
 - 3 days
 - Week 1
 - Week 2
 - Week 3
 - Week 4
 - Before Week 4 (everything)



Text meta: 85% accuracy on churn prediction



Accuracy: 85.2%
Precision: 76.9%
Recall: 66.7%

Train 60% | Test 40%

Stratified, 5 folds CV

NLP sentiment analysis

- VADER
- Off-the-shelf BERT



Sometimes I wonder why we **torture** you guys so badly 😢

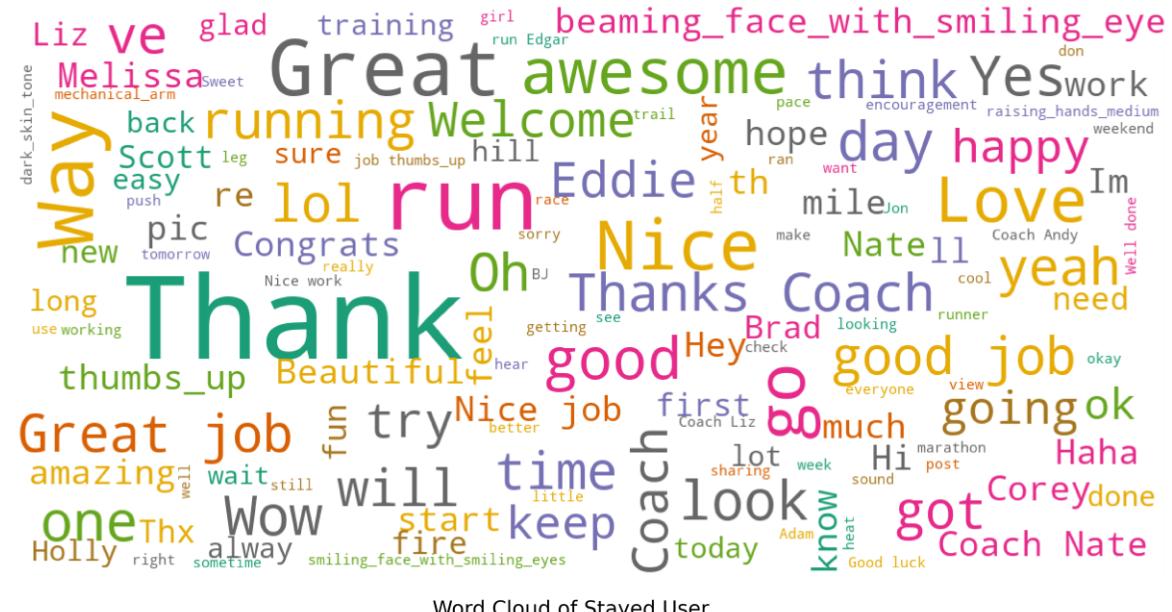
VADER: **-0.8356**

Off-the-shelf BERT: **negative 97%**

Overview of user texts

“Thanks Coach” “Good Job” “Great”

- Similar high frequency keywords
 - Positive & Supportive



Word Cloud of Stayed User



Word Cloud of Churned User

Fine-tuned BERT for sentiment analysis

BERT fine-tuned

Tone (positivity)

- Positive, neutral, ~~negative~~

Content (subjectivity)

- Rich, partial, none

Positive

“Congrats! Good job!”

- Content score: 0 (none)

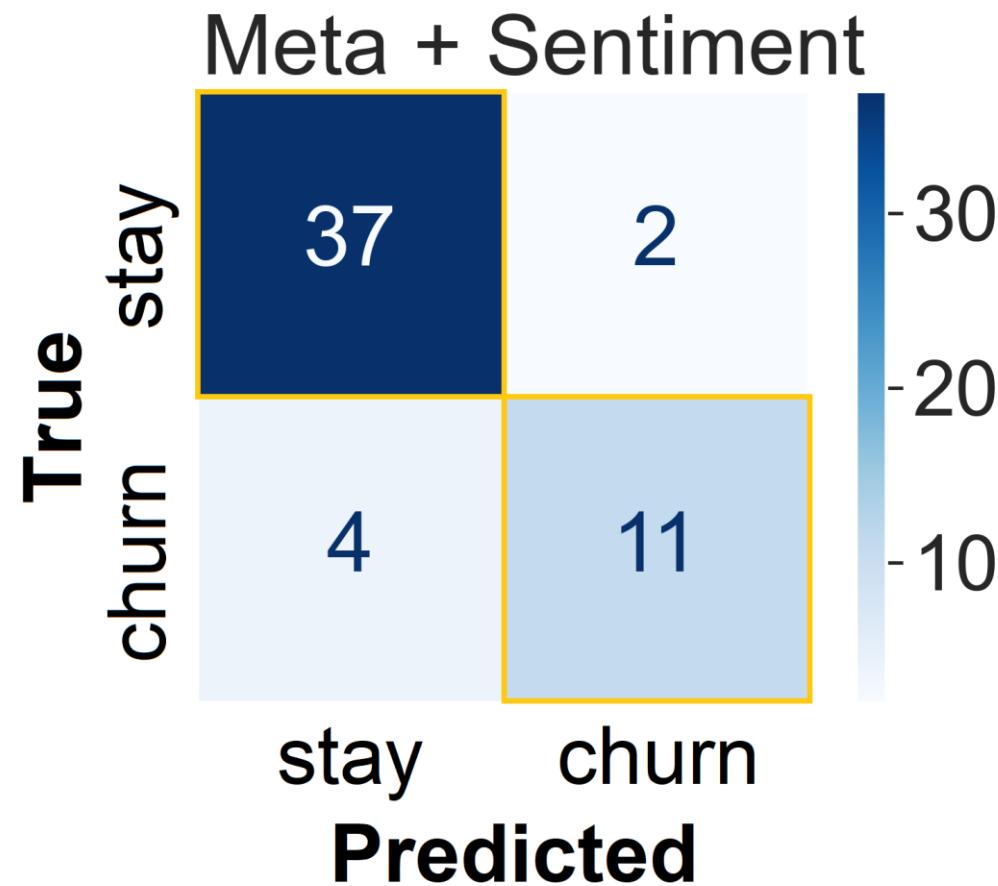
*“Congrats! Good job **on your first 10 miles!**”*

- Content score: 0.5 (partial)

*“Congrats! Good job **on your first 10 miles!** I had my first 10 miles this week too. It was BRUTAL cuz I had to do it in the full sun at the hottest part of the day. But I think it was REALLY good for me!”*

- Content score: 1 (rich)

Meta + Sentiment: 89% accuracy (Meta: 83%)

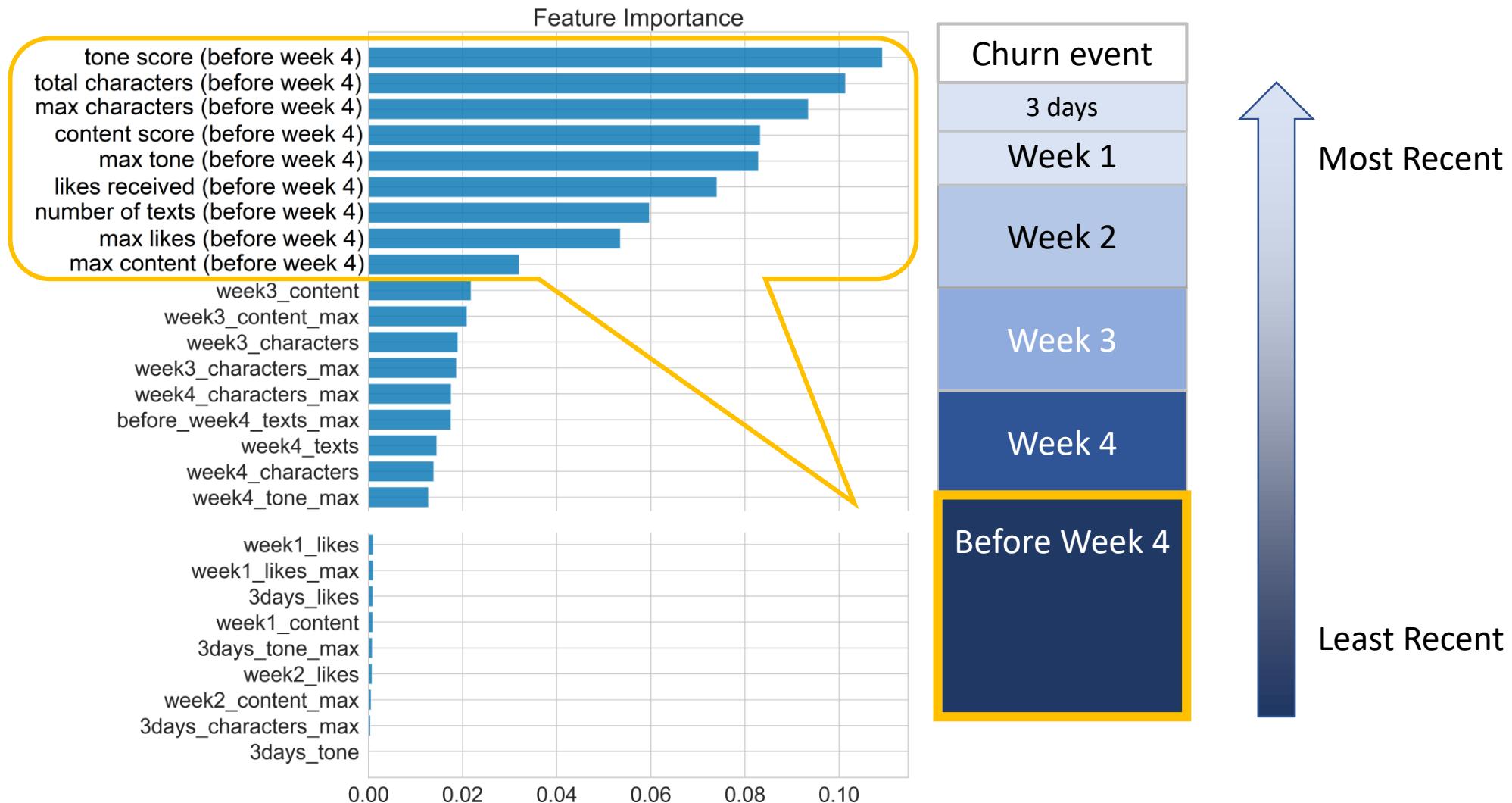


Accuracy: 88.9%
Precision: 84.6%
Recall: 73.3%

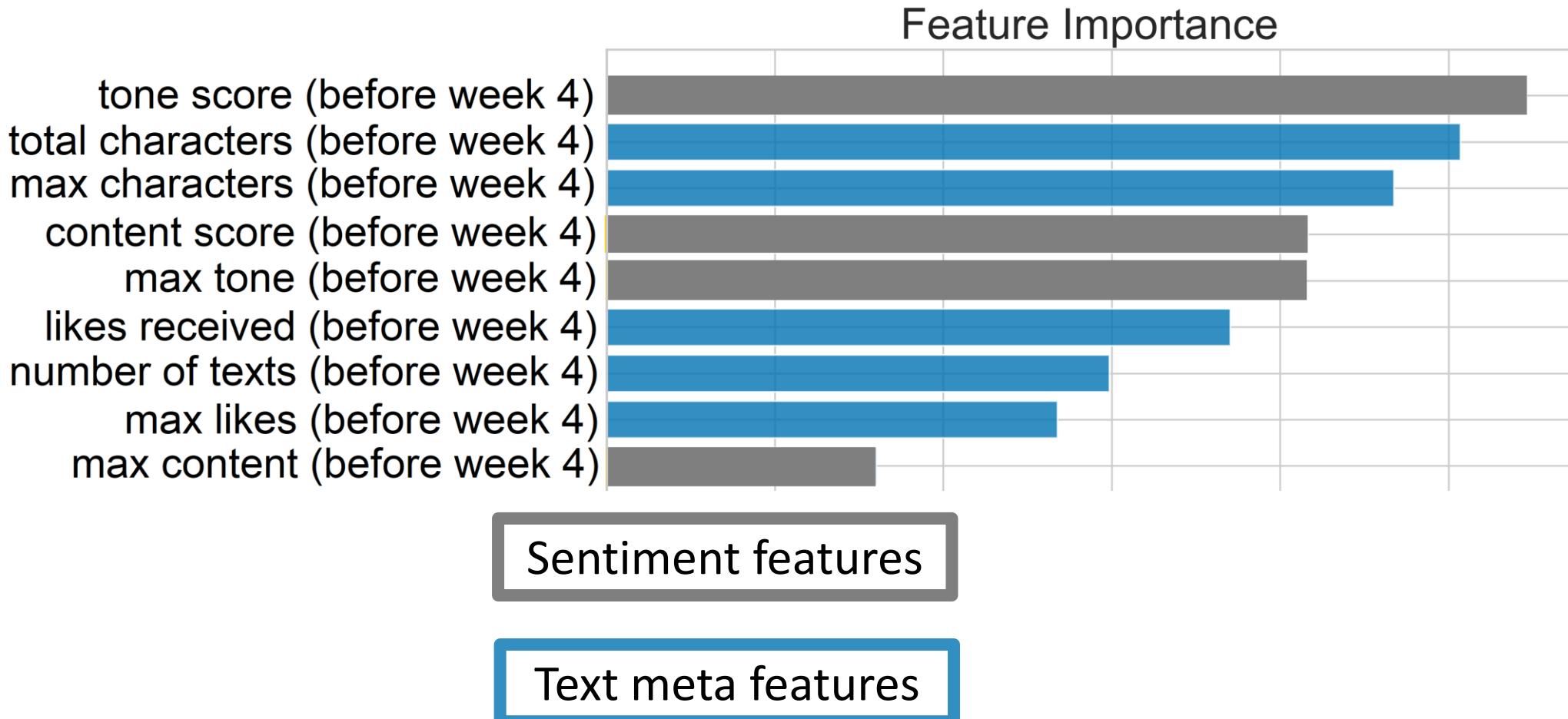
Train 60% Test 40%

Stratified, 5 folds CV

Top features, 4 weeks ago



Sentiment and meta features, comparable

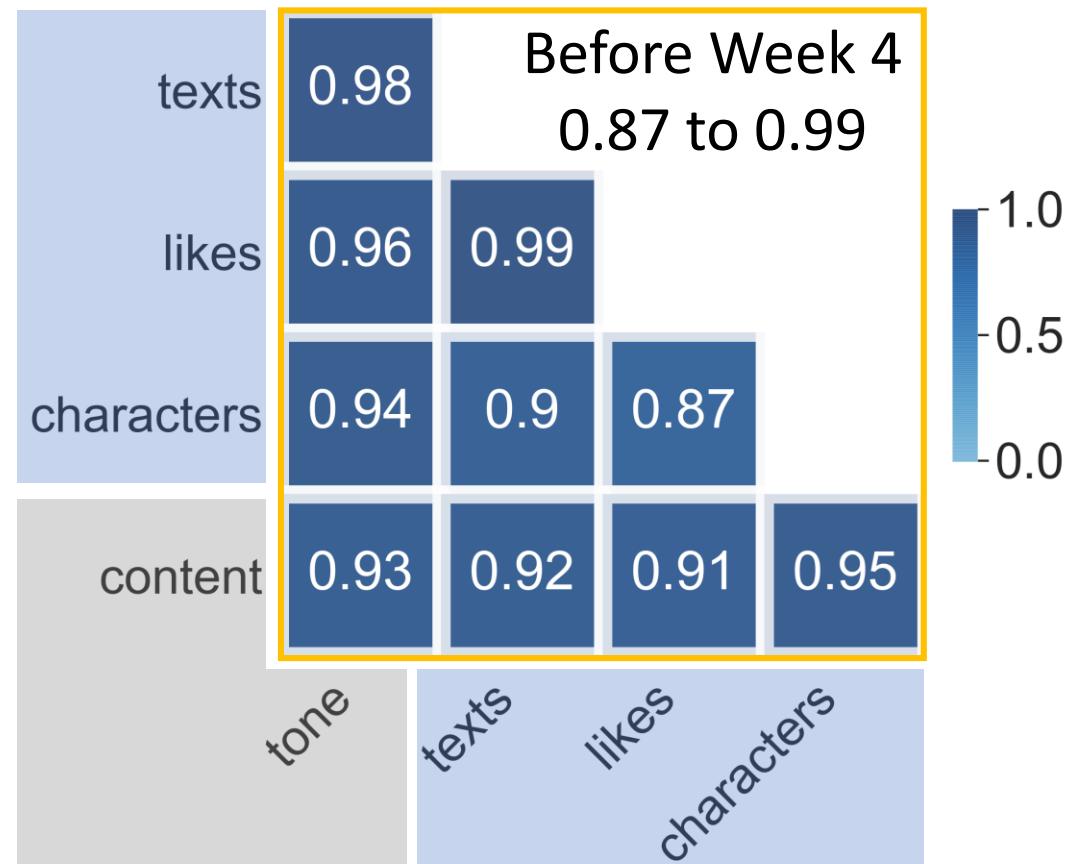


Strong correlation

Meta features

Sentiment features

- Why?
 - Quality vs Quantity



Text meta can predict user churn

Text data

- Accuracy 85% (meta)
- Accuracy 89% (meta + sentiment)

	Meta	Meta + Sentiment
Accuracy	0.85	0.89
Precision	0.77	0.84
Recall	0.67	0.73

Text meta good enough

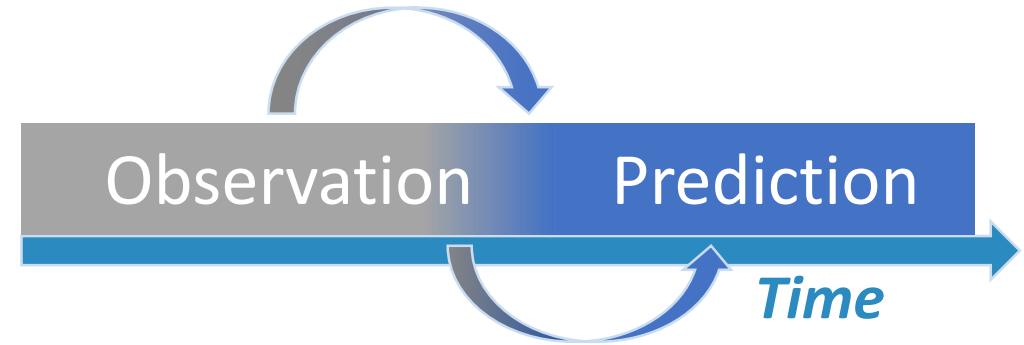
- Save time 50% (NLP labeling, transfer learning)
- Easy to scale up

Next

With more data

Real-time prediction

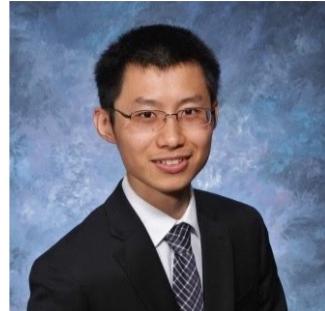
- Observation and prediction windows
- Feedback loop to update ML model



Topic analysis

- LDA / tf-idf
- User segmentation

Zelong (Eric) Zhang

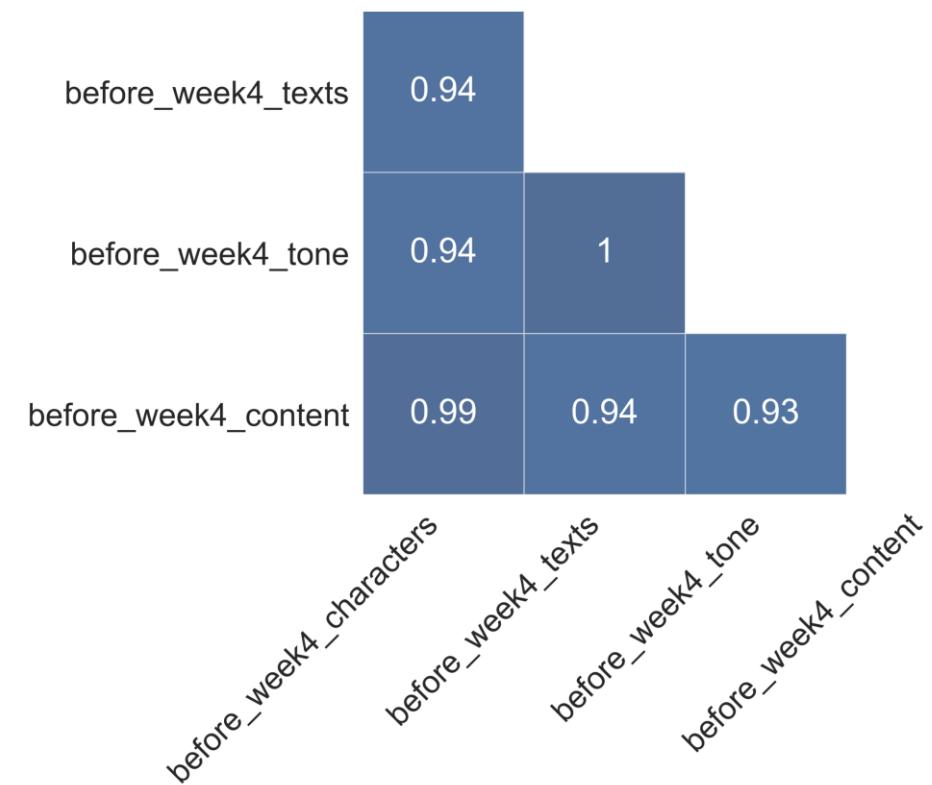


- PhD in Computational Chemistry
- Award-winning film (US DOE), photography
- User Experience and Decision-Making

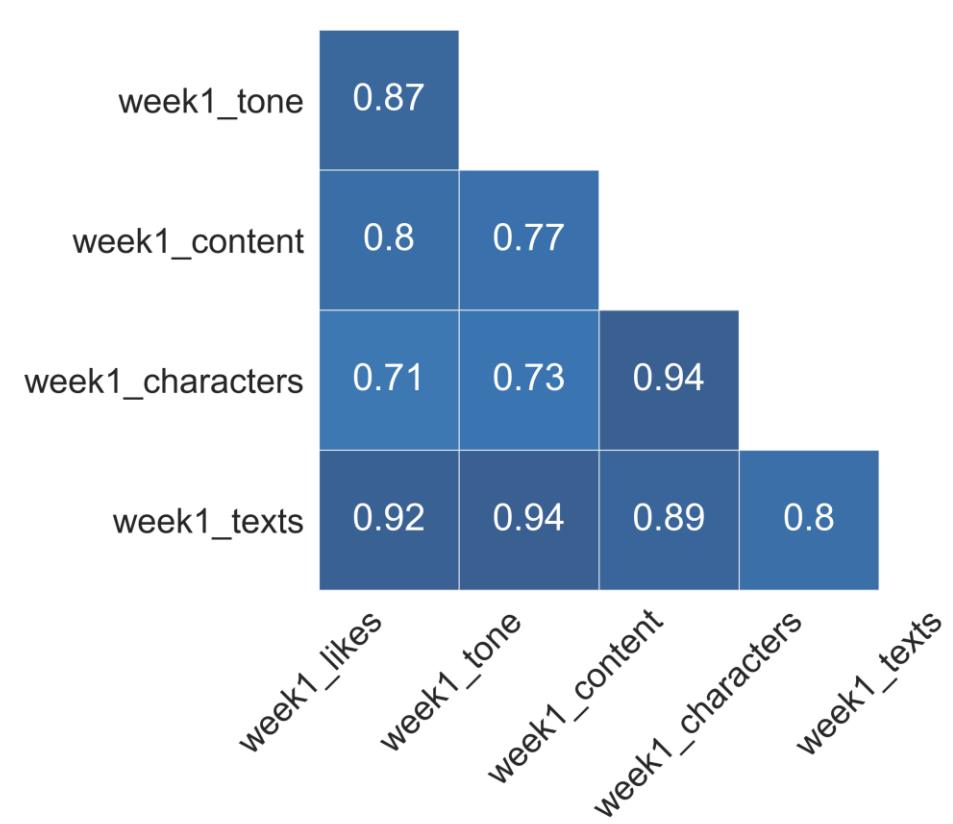


Extra Slides

before_week4_characters



week1_likes



1.0
0.5
0.0
-0.5
-1.0

Text meta-data can predict user churn

Strong correlation

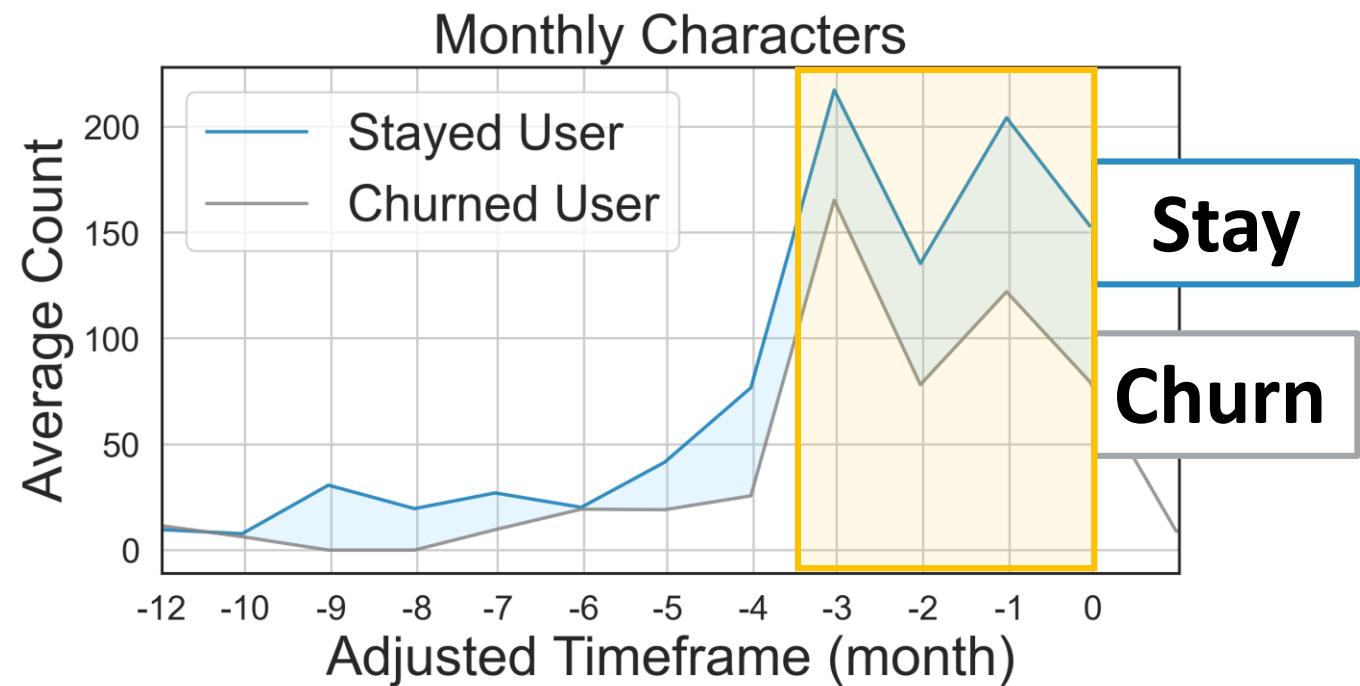
- text meta
- sentiment

Text meta features

- Good enough
- Easy to scale up

User in-app communication

- Strong indicator of user churn
- Customer life time 3 to 4 months

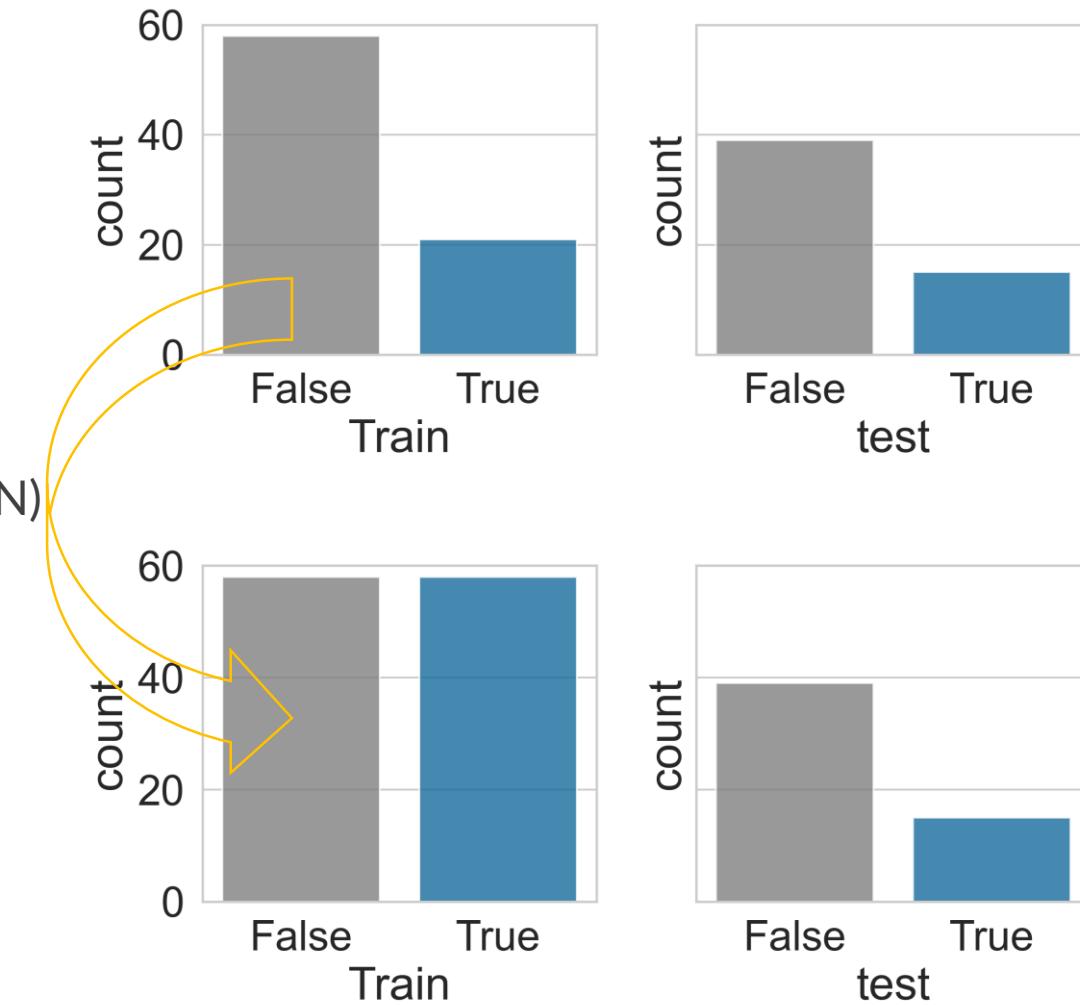


Train and test datasets

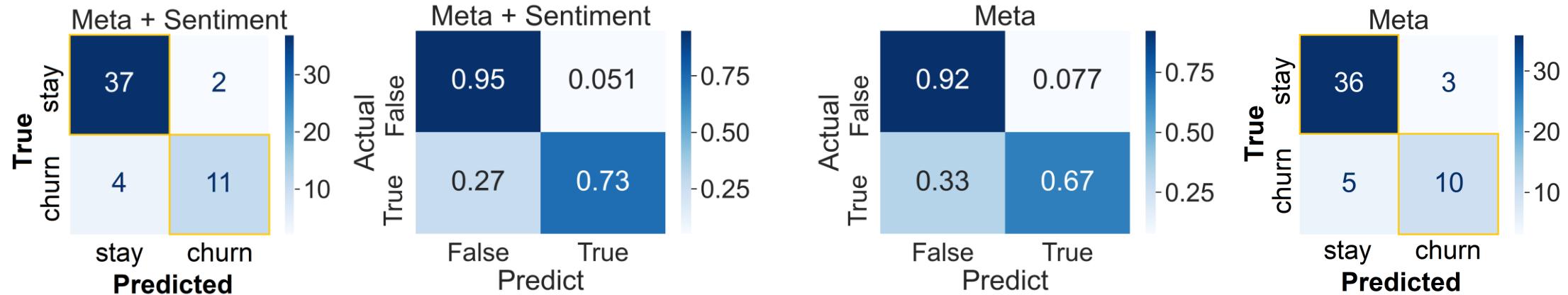
Adaptive Synthetic (ADASYN)

ADASYN is marginally better than SMOTE for this dataset (CV has slightly better recall).

"ADASYN focuses on generating samples next to the original samples which are wrongly classified using a k-Nearest Neighbors classifier while the basic implementation of SMOTE will not make any distinction between easy and hard samples to be classified using the nearest neighbors rule."



Validation metrics (cm: normalized)



	Meta + Sentiment	CV*		Meta	CV*
Accuracy	0.89	0.93 (0.035)	Accuracy	0.85	0.91 (0.049)
Precision	0.84	0.95 (0.040)	Precision	0.77	0.92 (0.047)
Recall	0.73	0.92 (0.074)	Recall	0.67	0.90 (0.097)
F1	0.79	0.93 (0.036)	F1	0.71	0.91 (0.054)

95% Confidence Interval of Accuracy (0.805, 0.973) vs (0.757, 0.947)

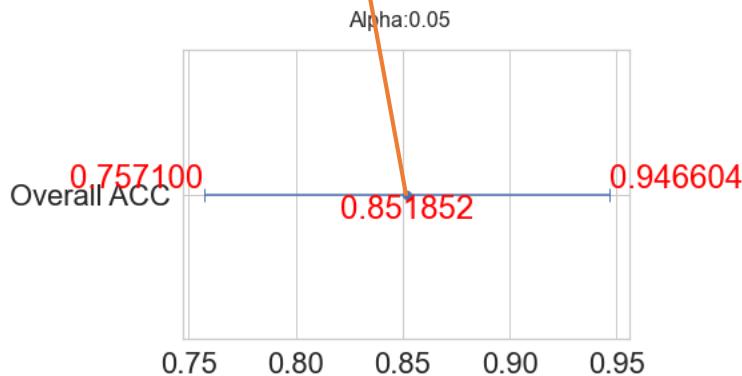
Cross-validation: stratified, 5 K-folds

95% Confidence Interval (accuracy, Kappa)

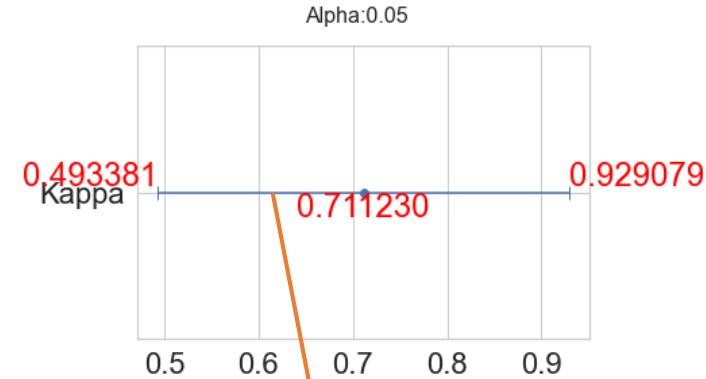
95%CI :
(0.8050661831398764, 0.9727115946379012)



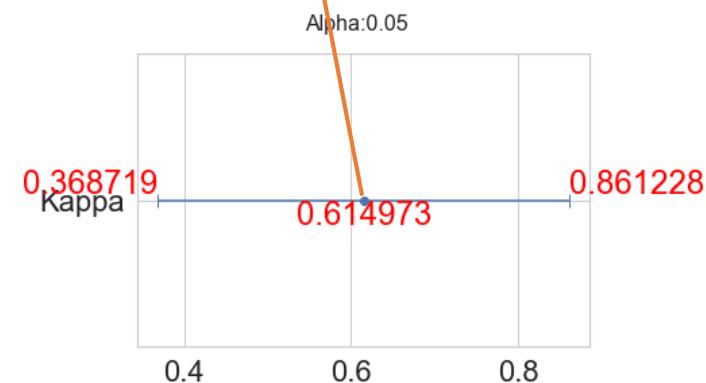
95%CI :
(0.757099643440483, 0.9466040602632207)



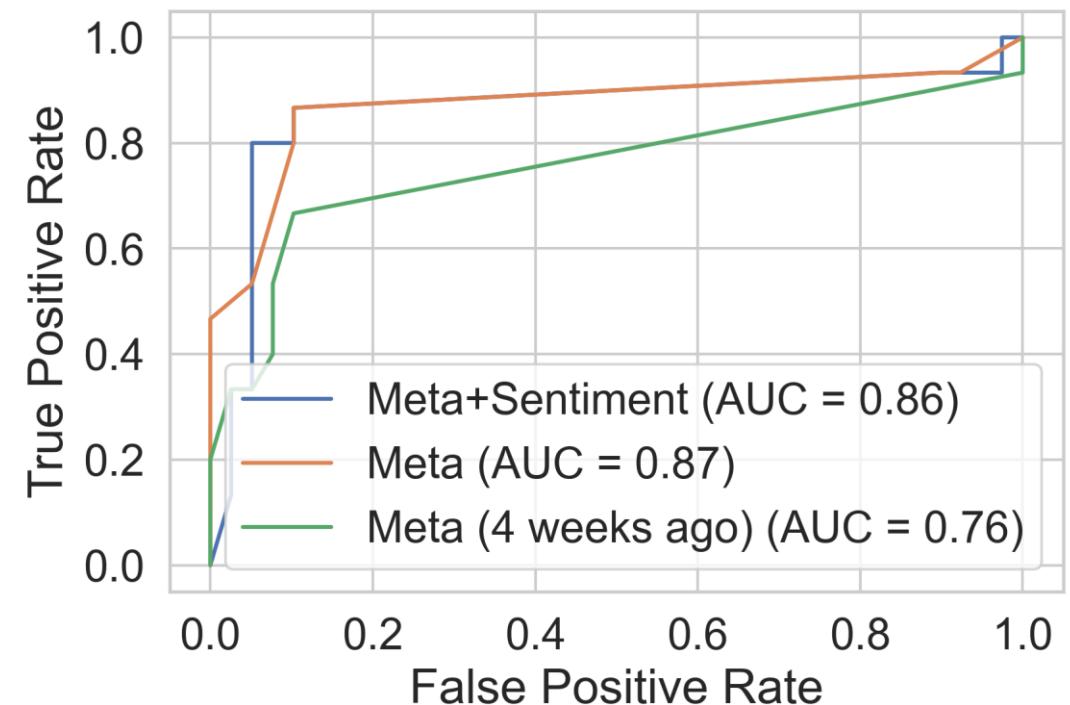
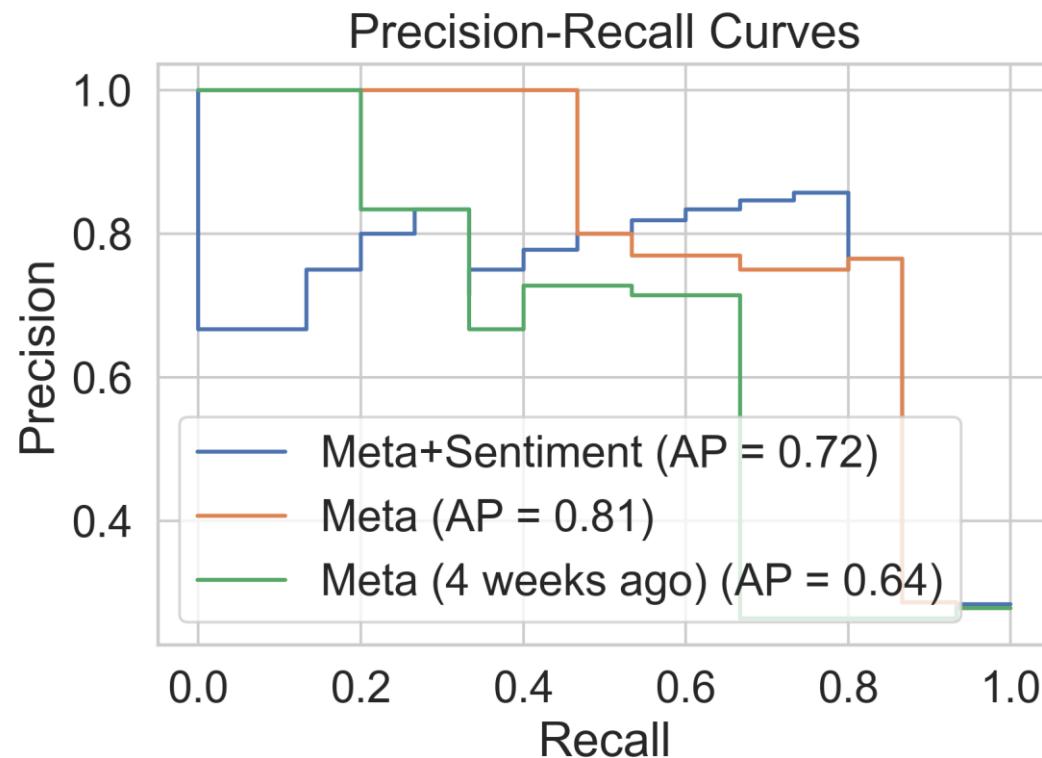
95%CI :
(0.49338056152930465, 0.9290793315188234)



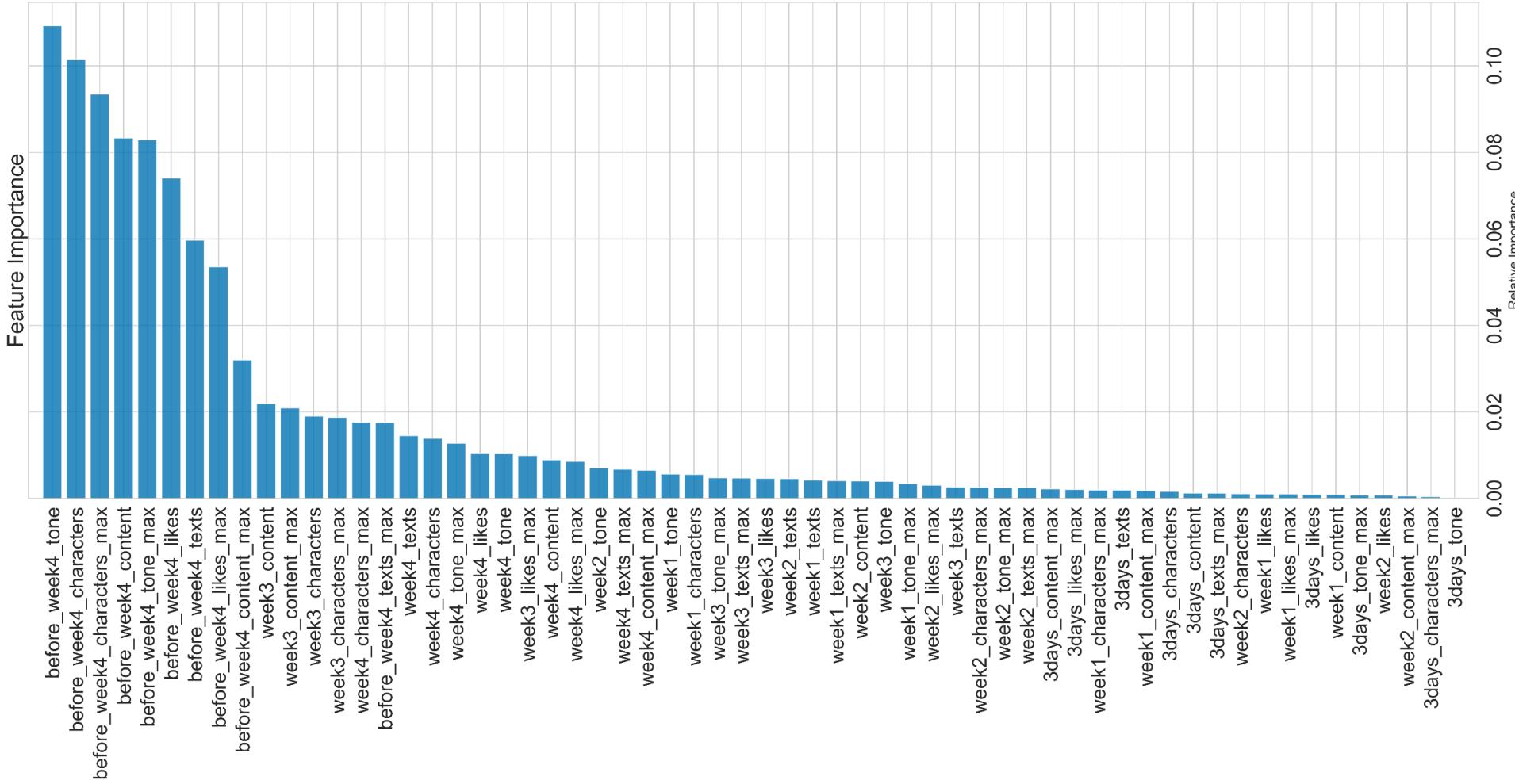
95%CI :
(0.36871885942285954, 0.8612276646413116)



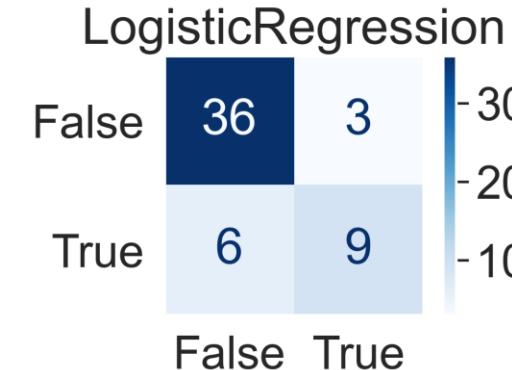
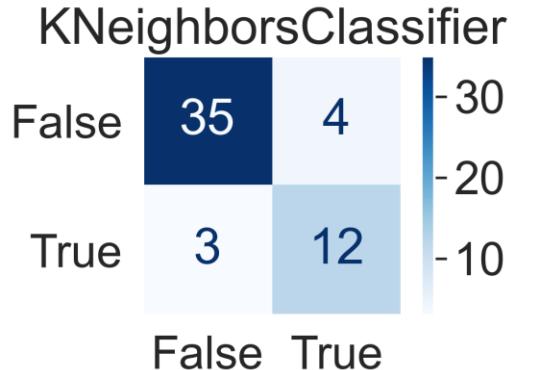
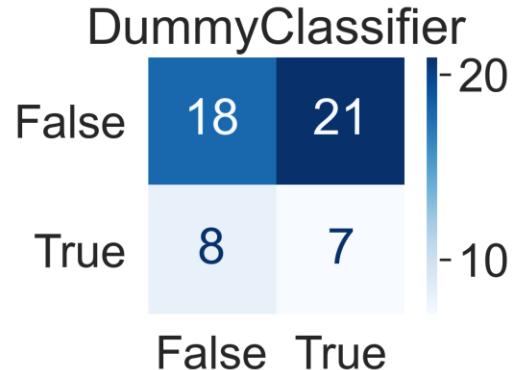
Precision-Recall curves and ROC curves



Feature importance

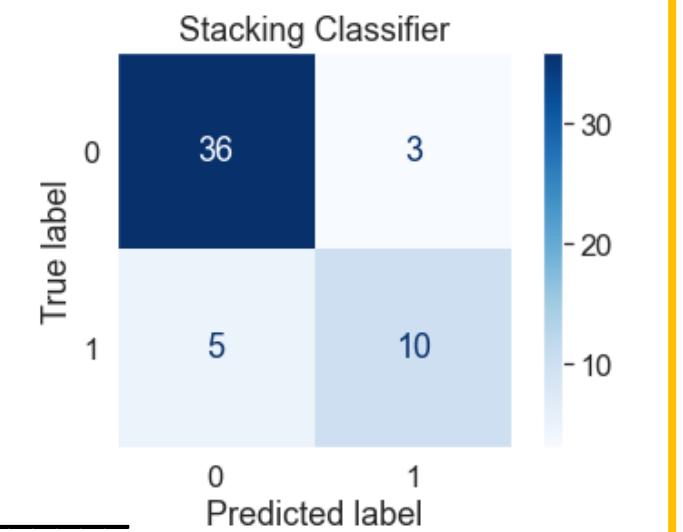
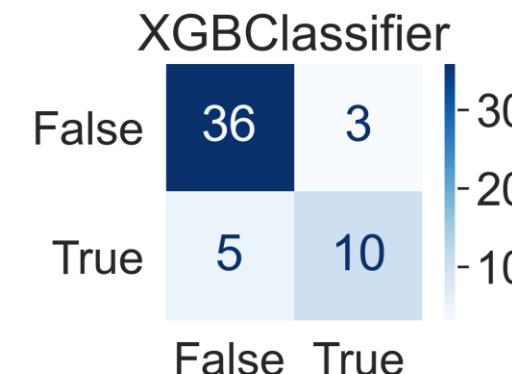
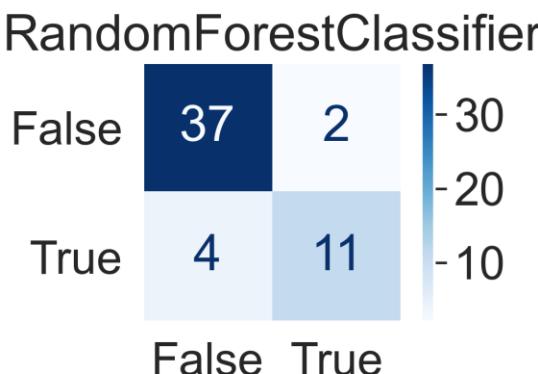
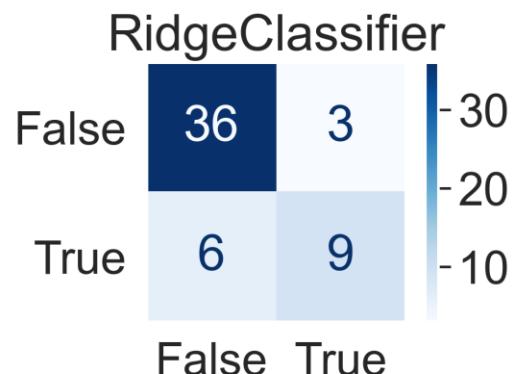


Classification models for churn



	Random Forest	Test	CV	Std. dev.
Accuracy	0.89	0.924	0.042	
Precision	0.846	0.951	0.040	
Recall	0.733	0.917	0.074	
F1	0.786	0.923	0.044	

CV: Stratified KFold, 5 splits

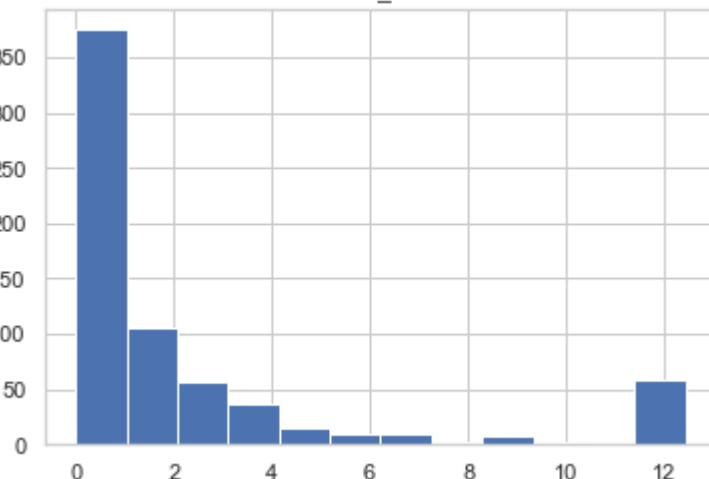


Combined by Logistic Regression

Churn User Lifetime Stats

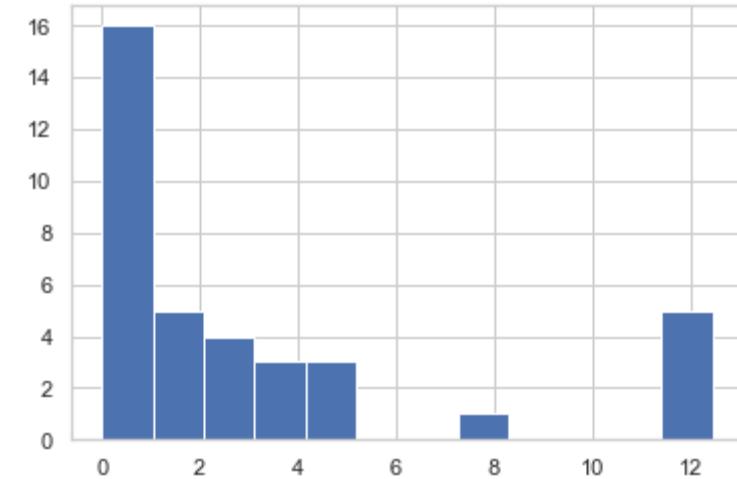
User Subscription	Month
Max	12.4
Min	0
Mean	2.8
Median	1.0
Unbiased variance	10.9
Standard deviation	3.3

lifetime_month

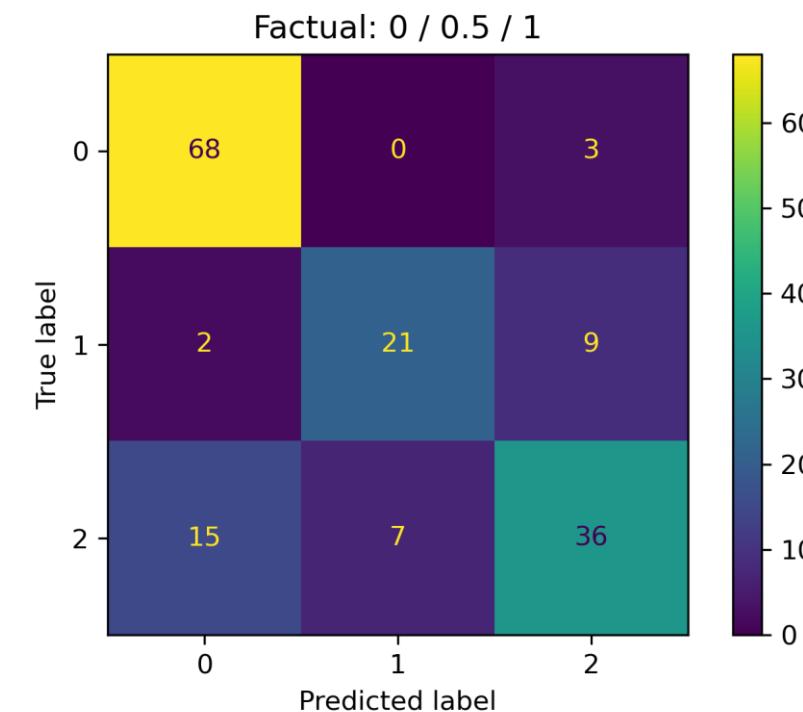
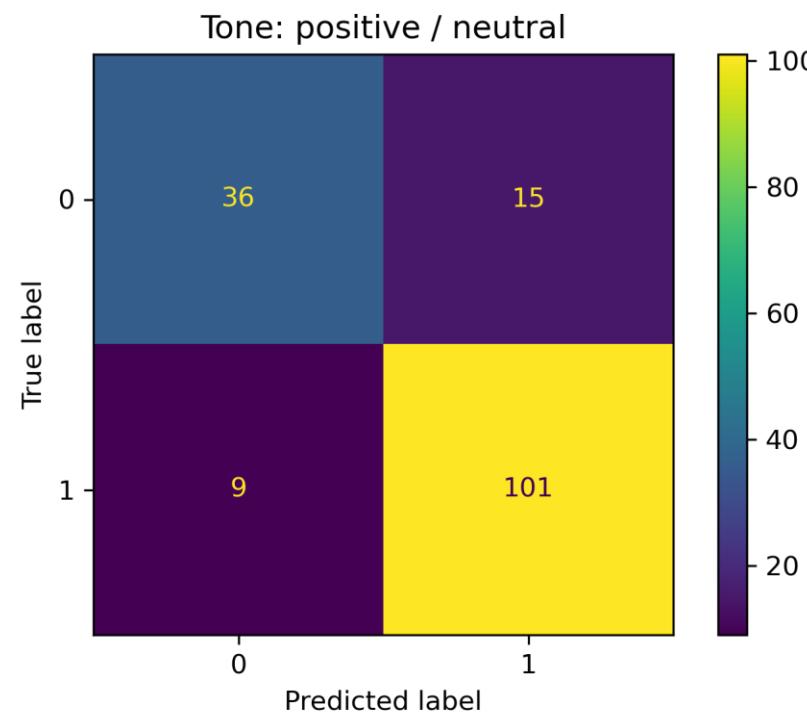


User in Text Data	Month
Max	12.4
Min	0
Mean	3.4
Median	2.0
Unbiased variance	15.2
Standard deviation	3.9

lifetime_m



NLP BERT validation metrics



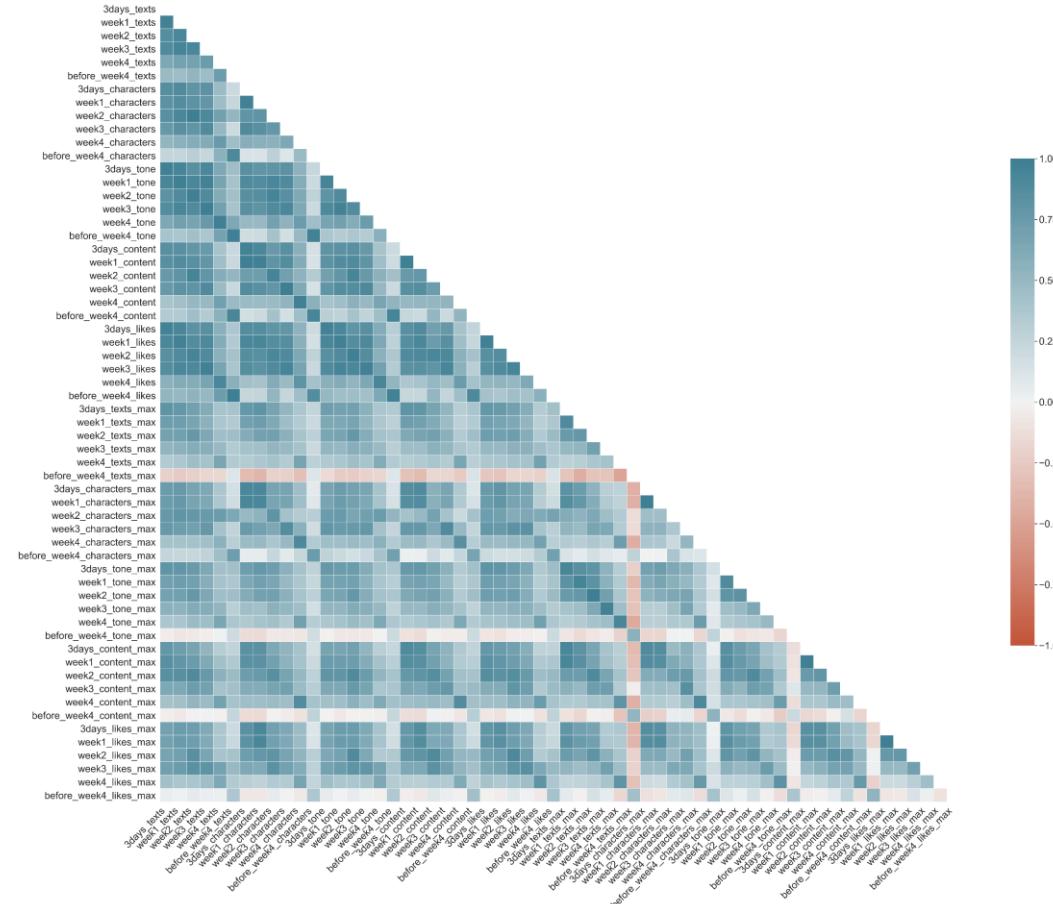
Tone:
accuracy 0.850932

Content:
accuracy 0.776398

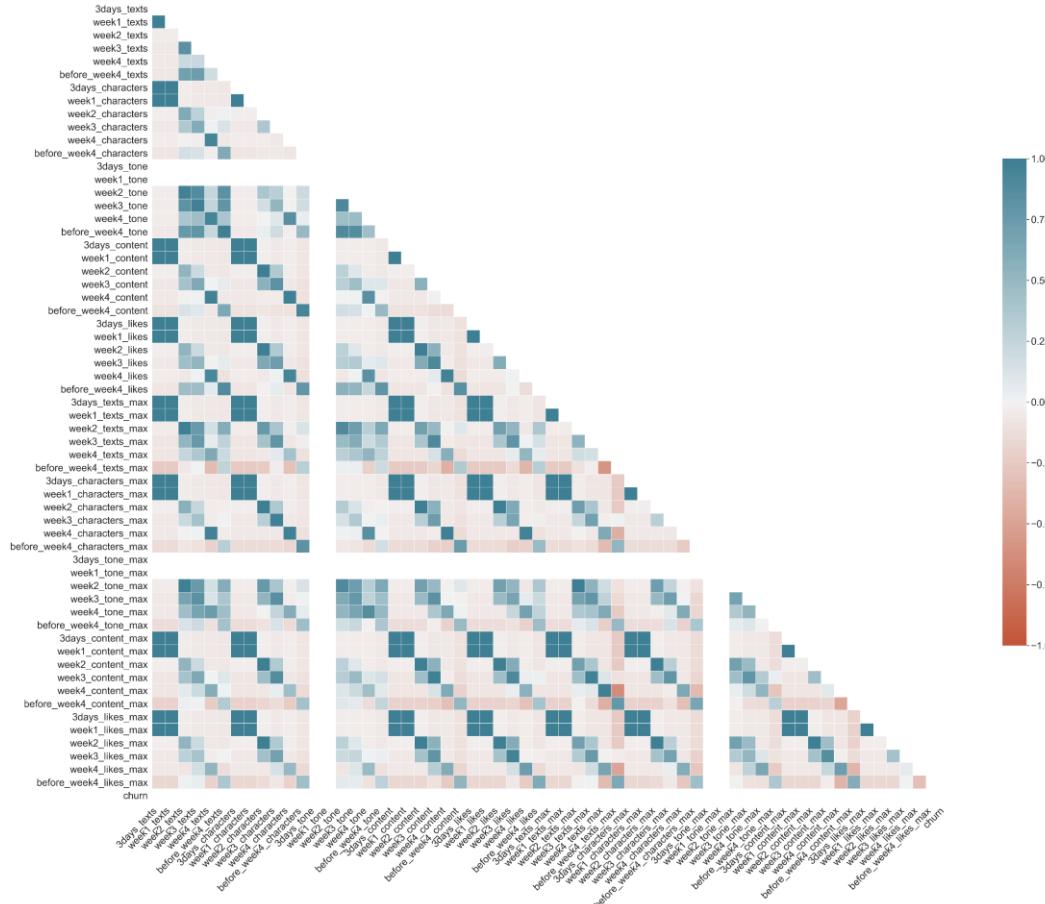
In-app user record

- Registered User: 56603
- User with text data: 773
- User with both registration and text data: 623
- User with subscription history: 2476
- User with both subscription history and text data: 159
- User with survey record: 490
- User with both subscription and survey: 124
- User with subscription, survey, and text: 34

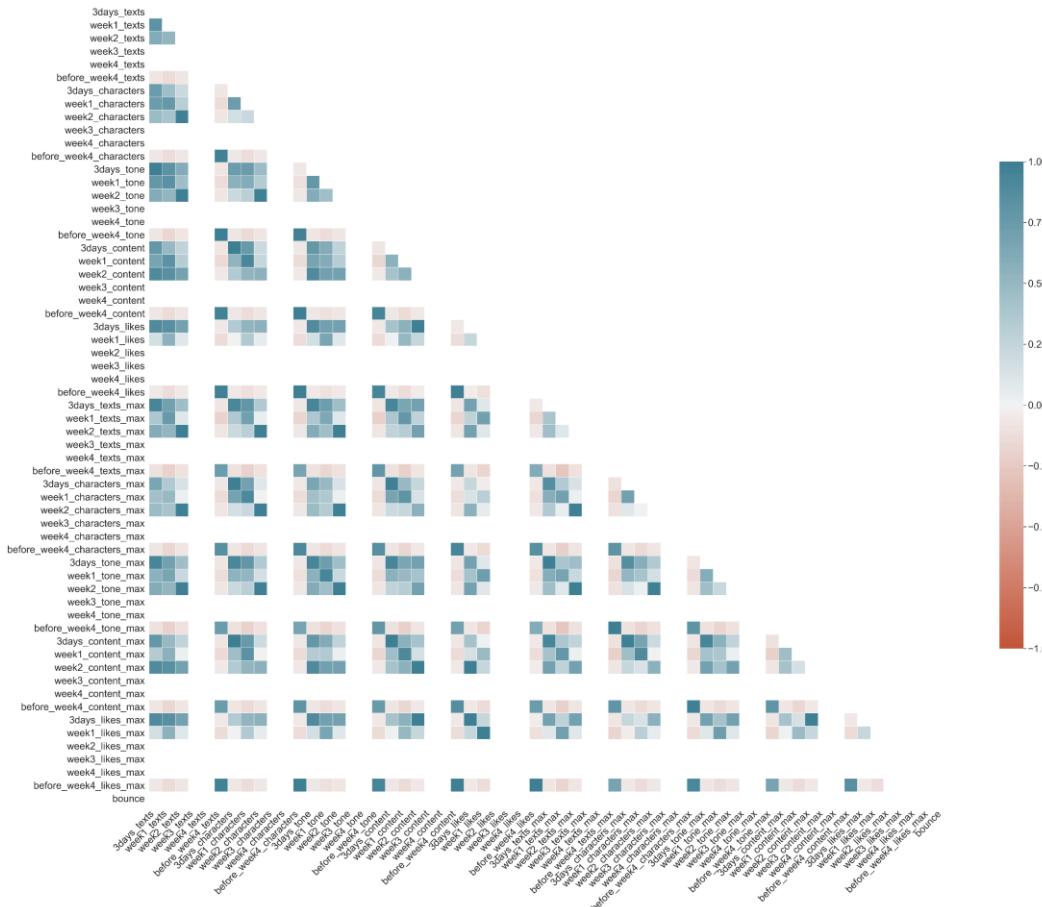
Feature correlation for active subscriber data

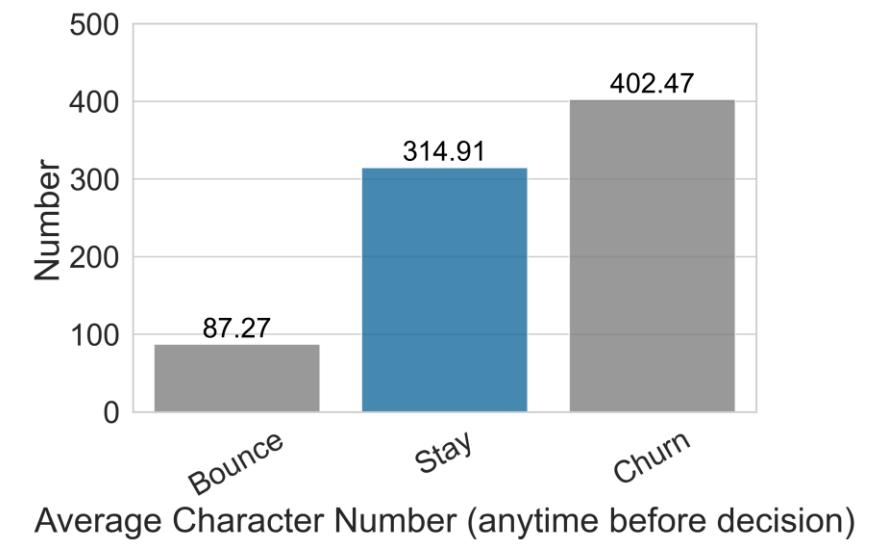
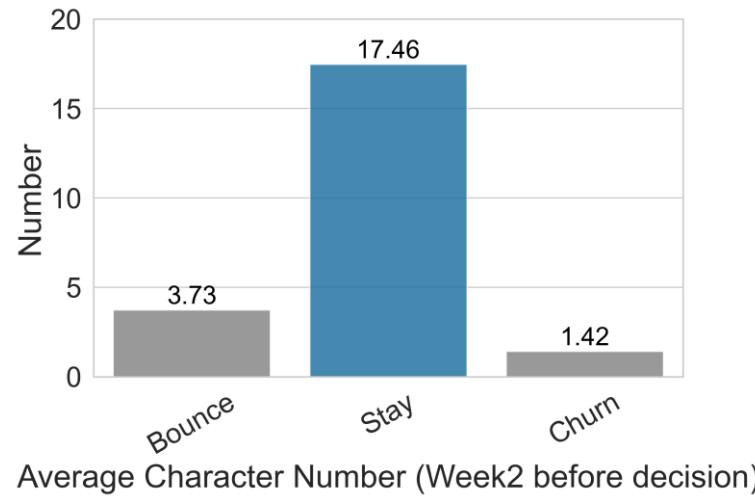
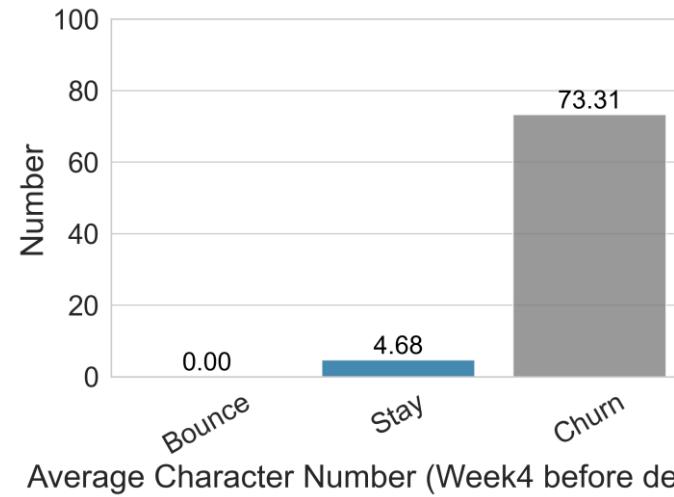


Feature correlation for churned user data

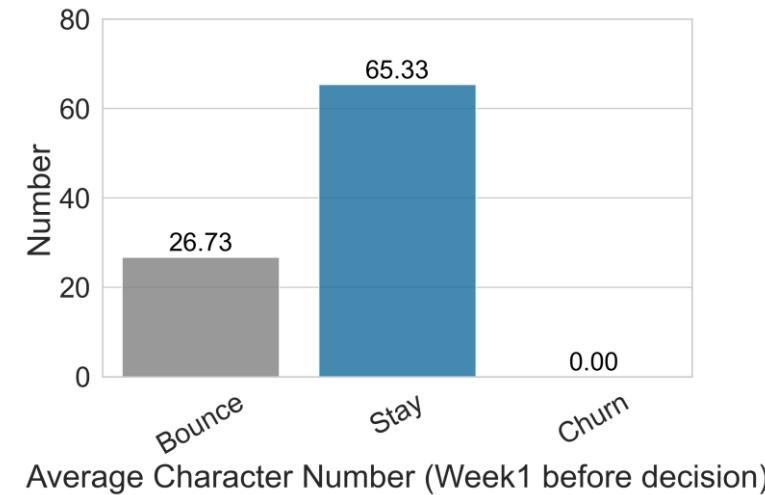
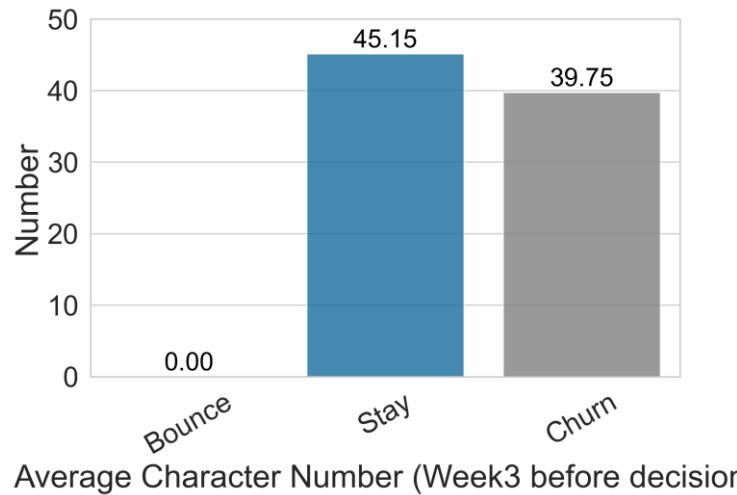


Feature correlation for bounded user data





Total character number averaged by user number



Time Series (original data)

