

Using content- and behavioural data for recommendations in the Norwegian news market

Bergens Tidende

Peter Kolbeinsen Klingenberg

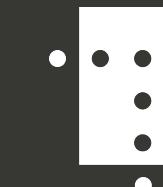
Research Centre for Responsible Media

Technology and Innovation

Project number 309339

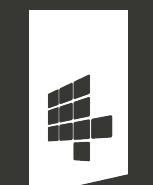


Norwegian Centre
for Research-based
Innovation



Nasjonalbiblioteket

NORCE



IBM

NRK

Schibsted

2

vizrt

amedia Bergens Tidende HIGHSOFT

Collaborators

- Bergens Tidende (BT)
- MediaFutures
- WP2 – User Modelling, Personalization & Engagement
- Supervisors
 - Mehdi Elahi (UiB/MediaFutures)
 - Thomas Husken (BT)

Bergens Tidende

Outline

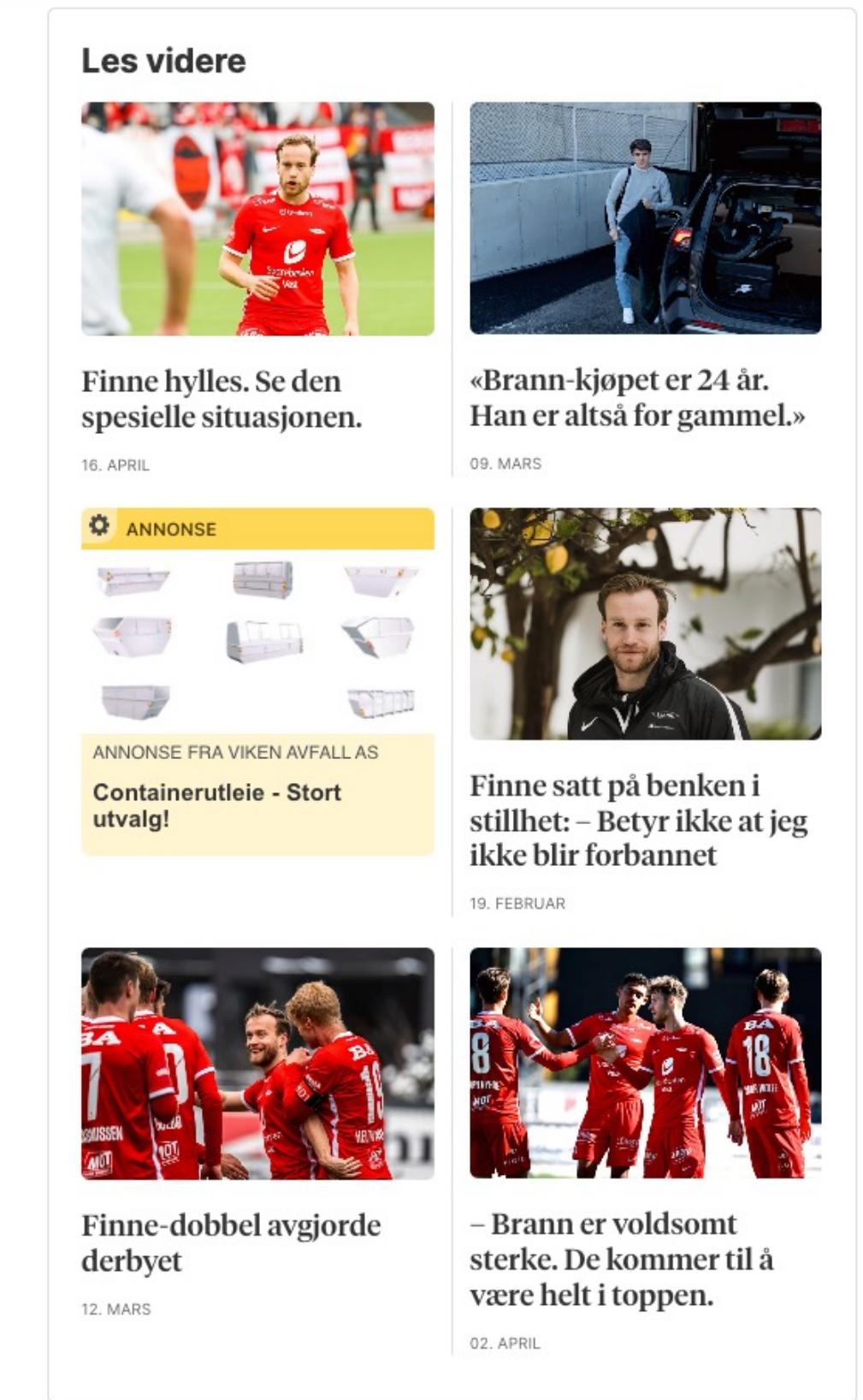
- Offline experiment
 - t-SNE dimension reduction of embeddings by **BERT**, trained on the Norwegian language
 - Machine learning models (**ALS**, **BPR** and **LMF**) trained on observed user interactions
- Online experiment
 - **A/B test** of **BERT** embeddings based on articles textual content, tested against a former model in BT based on **generic tags**

Why provide recommendations?

- **Identify and automate** user interest/preference
- Satisfying recommendation
 - Users **spending more time**
 - Gain **more activity**
- Previous articles **more accessible**

The research

- The research concerns
 - How to effectively model the reading behaviours of the users in Norwegian media platforms and the content of news articles in the Norwegian language?
 - Which technique can be used to generate personalised recommendations for the users incorporating these two types of data, i.e., the users behaviour and the content of news?
- “Most Popular”, “Most Debated”, “More Articles” and “Read Further”



Datasets

- Timespan
 - 12.09.22 – 07.11.22
- Content data about articles
 - 59 features
 - 10 134 articles
- User behavioural data
 - 48 227 articles
 - 18 100 663 interactions/clicks
 - 179 761 users

Recommendation of content data

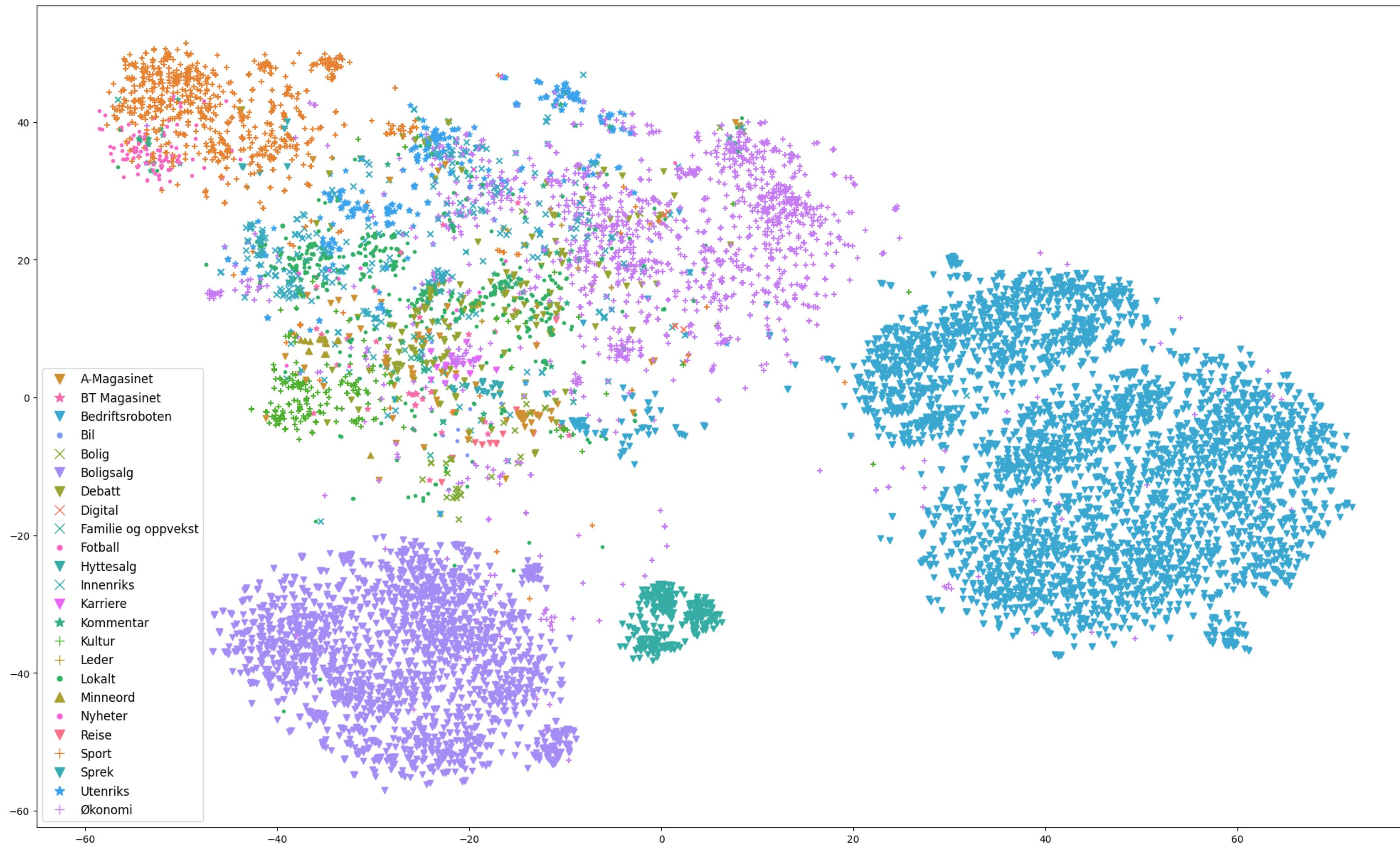
- Sentence embeddings
 - **Numerical dense vector** representation of sentences
- National Library of Norway's AI Lab
 - **NB-SBERT-BASE**
 - November 2022
 - Bokmål and nynorsk
 - Digital collection from National Library
 - Texts from the last 200 years

Embeddings and how it is used

- 768 elements/dimensions
- Represents the **semantic content** of a sequence of text
- Cosine similarity
- Value between 0-1
- Iterates through a set of other embeddings of articles
- Fetch top 6

```
array([-0.08294732, -0.28578746,  0.45302293,  1.0732058 , -0.05859034,
       -0.09161276,  0.25284255, -0.83497876,  0.21518148,  0.18443395,
       0.94029593, -0.35844457, -0.31025425, -0.5126164 ,  0.89932996,
       0.01906117,  0.10264873,  0.30682546, -0.01190989,  0.03831028,
      -0.6800012 ,  0.45138714, -0.7301062 , -0.24777927, -0.1340554 ,
      -0.2192291 ,  0.09935746, -0.13037616,  0.17330395, -0.23620069,
     -0.10453023,  0.45933157,  0.47540715, -0.13308021, -0.17357284,
      0.2189701 ,  0.50240254, -0.322204 ,  0.235362 ,  0.37498638,
      0.00888087, -0.6357252 ,  0.49922812,  0.53972644, -0.34229192,
      0.51335466, -0.45986283,  0.14874454, -0.6273218 , -0.22660296,
      0.1271185 , -0.43681327,  0.17014284, -0.19972932,  0.13008371,
      0.97933424, -0.11915381,  0.46735737,  0.18857948, -0.29300576,
     -0.03847165,  0.34857717,  0.6889674 , -0.43690094,  0.47072905,
      0.22244413, -0.1462297 , -0.4992958 ,  0.29579484, -0.8806289 ,
     -0.33545762, -0.559921 ,  0.14002588,  0.60932624, -0.5361552 ,
     -0.44619134,  0.04705823,  0.64950883,  0.26500487, -0.13775769,
      0.04189495, -0.13227613, -0.2425658 ,  0.4876327 , -0.19441366,
     -0.05705135, -0.32327873, -1.0409969 ,  0.15237664, -0.01390435,
      0.07377632, -0.7382443 ,  0.19637907, -0.87199146, -0.39569667,
     -0.09288712,  0.2245904 ,  0.3941891 ,  0.3106181 ,  0.11970903,
     -0.29882726, -0.3370344 , -1.0512552 ,  0.02632958,  0.07893211,
     -0.11072347, -0.03400832, -0.55372226,  0.40065378,  0.1149893 ,
     -0.43493524,  0.06644171, -0.1497159 ,  0.13212778,  1.163495 ,
     -0.14863054,  0.16408837,  0.18070263, -0.77492946, -0.03854357,
      0.43890914, -0.24989437,  0.5815154 ,  0.7047419 , -0.26090434,
      ...
      -1.2491299 , -0.5011168 , -0.10211495, -0.35821217, -0.17122957,
     -0.5984507 , -0.7049129 ,  0.14699967,  0.23942627, -0.07543122,
     -0.22193967, -0.29848447,  0.6426493 ,  0.5291974 ,  0.04088476,
     -0.08763585,  0.16316445, -0.1757913 , -0.71188855, -0.08565662,
     -0.22070274,  0.3358265 ,  0.17543437])
```

t-SNE dimension reduction on all articles



Results of dimension reduction

- Ability to cluster categories quite well
- Manually checks
- No exact performance measures

Recommendation of user behaviour data

- Train/test split (80/20)
- Split on users
 - Trained on all articles, **80%** users
- Machine learning models
 - **ALS** (Alternating Least Squares)
 - **BPR** (Bayesian Personalized Ranking)
 - **LMF** (Logistic Matrix Factorization)
- Baseline models
 - **Most Popular**
 - **Random**

Evaluation metrics

- Precision@K
- Recall@K
- MAP@K (MeanAveragePrecision)
- NDCG@K (NormalizedDiscountedCumulativeGain)
- Area Under the ROC Curve (Reciever Operator Characteristics)
- Area Under the Precision-Recall Curve

Experiments of user behaviours

- Filtering stages
 - F1 – Remove all clicks **outside timespan**
 - F1 – Remove all clicks on “**Direkte**” section in BT
 - F2 – Articles has to be read by **two or more** to be included
 - F2 – Every user is required to **read one** article per week
- Hyperparameter tuning of **ALS**, **BPR** and **LMF** including filter F1 and F2

Evaluation F1

- Applied general filters
- Around **1,5 million** interactions, **18k** users and **40k** articles excluded
- Default hyperparameters used

| Dataset | Interactions | Users | Items |
|----------------|--------------|---------|-------|
| User Behaviour | 16 657 690 | 161 092 | 7 769 |

| Algorithms | Evaluation metrics | | | | | |
|--------------|--------------------|---------------|---------------|---------------|---------------|---------------|
| | P@5 | R@5 | MAP@5 | NDCG@5 | ROC_AUC | PR_AUC |
| Random | 0.0046 | 0.0005 | 0.0021 | 0.0046 | 0.5006 | 0.0061 |
| Most Popular | 0.1679 | 0.0295 | 0.1276 | 0.1829 | 0.9326 | 0.0977 |
| ALS | 0.0868 | 0.0135 | 0.0566 | 0.0927 | 0.6168 | 0.0416 |
| BPR | 0.0721 | 0.0139 | 0.0470 | 0.0773 | 0.6392 | 0.0300 |
| LMF | 0.0271 | 0.0080 | 0.0152 | 0.0289 | 0.8685 | 0.0298 |

Evaluation F2

- Applied specific filters, in addition to general filters from F1
- Additionally, **1 million interactions, 90k users and 1k articles are excluded**
- Default hyperparameters used

| Dataset | Interactions | Users | Items |
|----------------|--------------|--------|-------|
| User Behaviour | 15 531 878 | 69 762 | 6 324 |

| Algorithms | Evaluation metrics | | | | | |
|--------------|--------------------|---------------|---------------|---------------|---------------|---------------|
| | P@5 | R@5 | MAP@5 | NDCG@5 | ROC_AUC | PR_AUC |
| Random | 0.0086 | 0.0006 | 0.0038 | 0.0084 | 0.5012 | 0.0120 |
| Most Popular | 0.2845 | 0.0220 | 0.2037 | 0.2950 | 0.9293 | 0.1388 |
| ALS | 0.1789 | 0.0136 | 0.1134 | 0.1825 | 0.7495 | 0.0839 |
| BPR | 0.1135 | 0.0086 | 0.0698 | 0.1183 | 0.6356 | 0.0406 |
| LMF | 0.0504 | 0.0035 | 0.0258 | 0.0500 | 0.8601 | 0.0491 |

Evaluation hyperparameter tuning

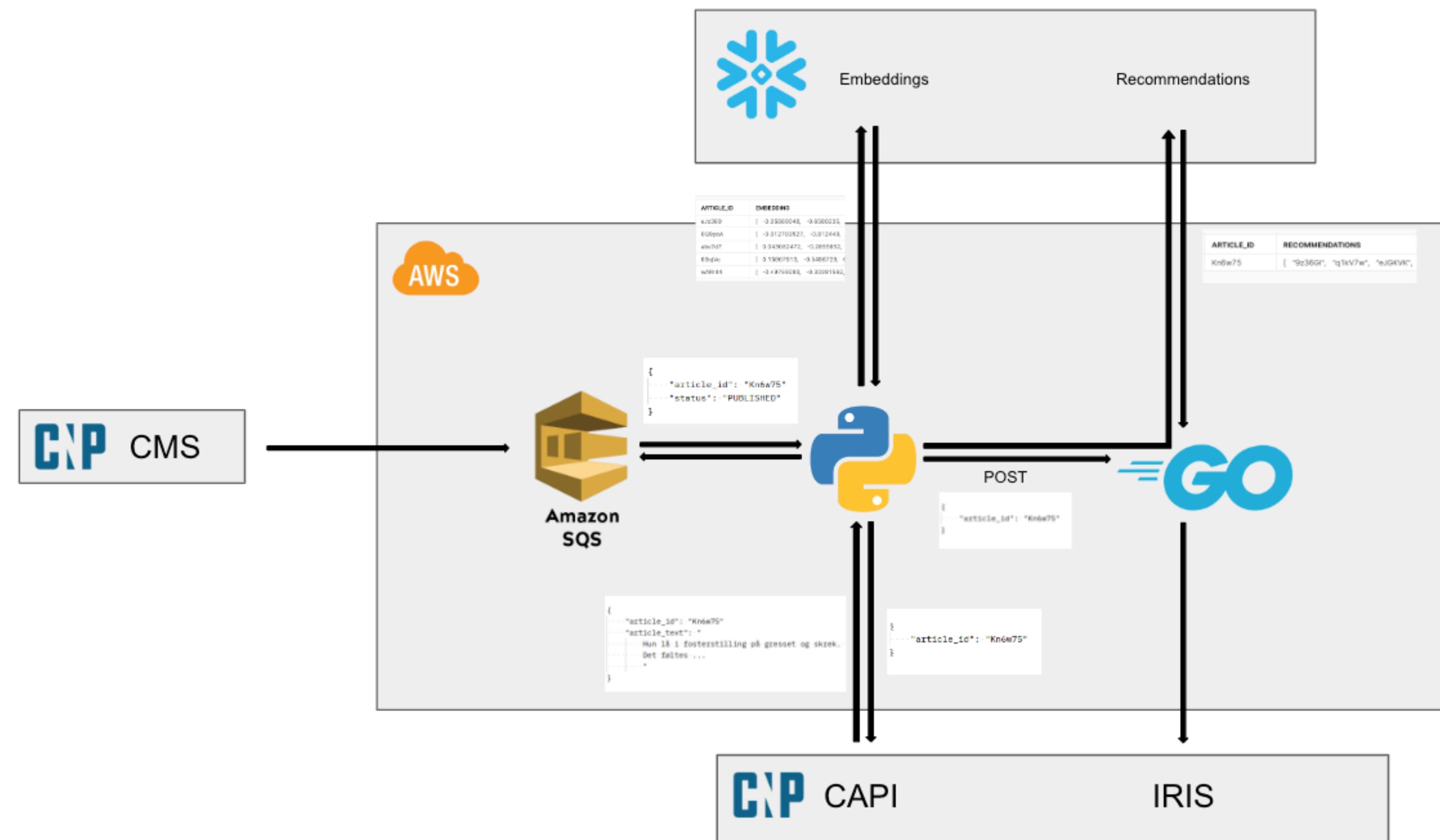
- Applied specific filters from F2, in addition to general filters from F1
- Factors, iterations, regularization and learning rate
- Best hyperparameters based on best P@5 value

| Dataset | Interactions | Users | Items |
|----------------|--------------|--------|-------|
| User Behaviour | 15 531 878 | 69 762 | 6 324 |

| Algorithms | Hyperparameters | | | |
|------------|-----------------|------------|----------------|---------------|
| | factors | iterations | regularization | learning_rate |
| ALS | 5 | 5 | 0.01 | N/A |
| BPR | 30 | 5 | 0.1 | 0.001 |
| LMF | 5 | 100 | 1.0 | 1.0 |

| Algorithms | Evaluation metrics | | | | | |
|--------------|--------------------|---------------|---------------|---------------|---------------|---------------|
| | P@5 | R@5 | MAP@5 | NDCG@5 | ROC_AUC | PR_AUC |
| Random | 0.0086 | 0.0006 | 0.0038 | 0.0084 | 0.5012 | 0.0120 |
| Most Popular | 0.2845 | 0.0220 | 0.2037 | 0.2950 | 0.9293 | 0.1388 |
| ALS | 0.2304 | 0.0178 | 0.1590 | 0.2381 | 0.8826 | 0.1166 |
| BPR | 0.2869 | 0.0222 | 0.2049 | 0.2966 | 0.6266 | 0.1028 |
| LMF | 0.1285 | 0.0092 | 0.0755 | 0.1295 | 0.8918 | 0.0720 |

BERT recommender architecture in BT



A/B test

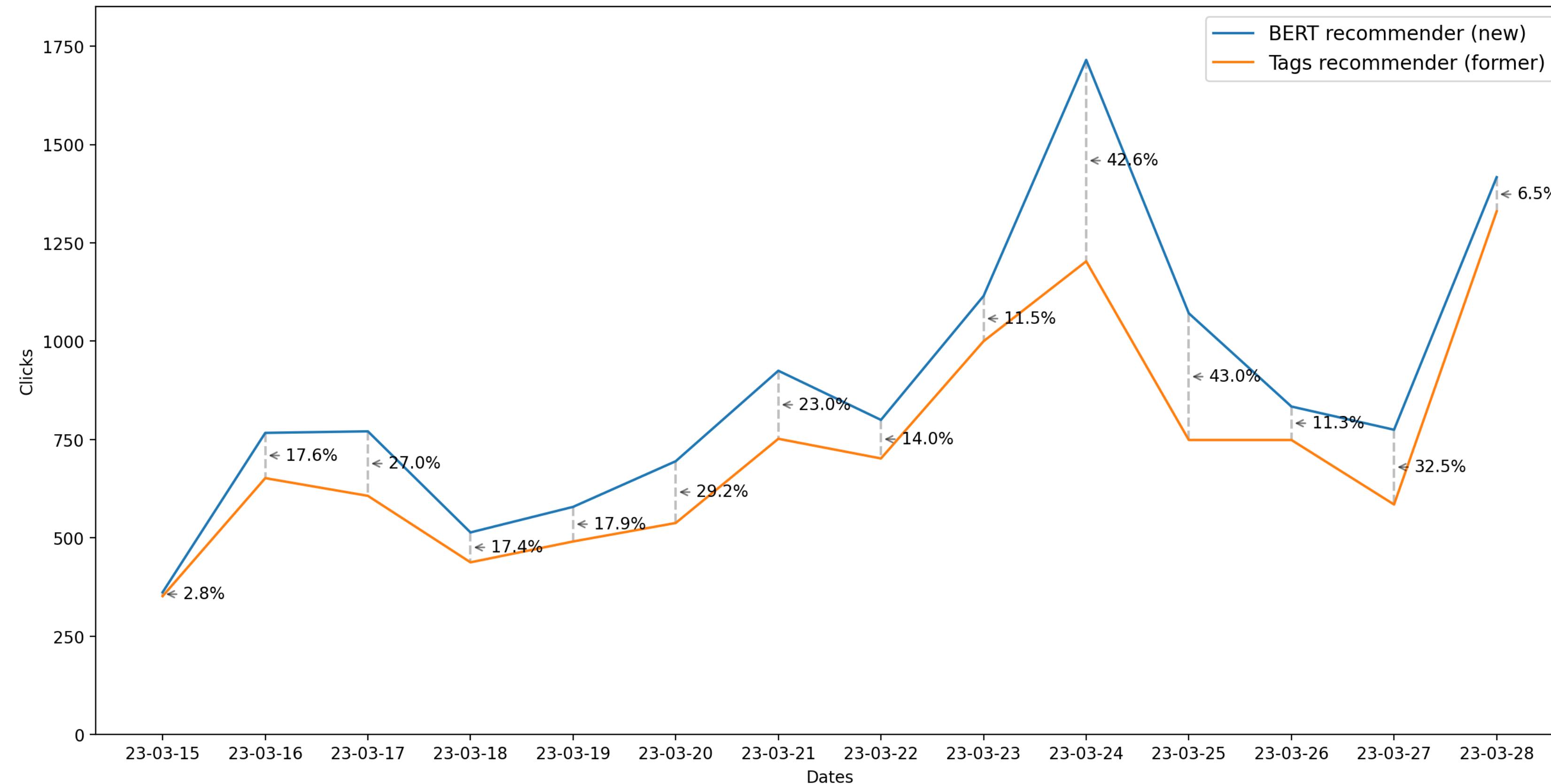
- **BERT** model tested against a **former recommendation** model in BT based on generic tags
- **50/50 split**
- Recommendation list is displayed at the **bottom** when reading an article

The screenshot shows a news article from Bergens Tidende. The article title is "Les videre" (Read more) and the subtitle is "Finne hylles. Se den spesielle situasjonen." (Find the shelves. See the special situation.). The date is 16. APRIL. Below the article, there is a yellow advertisement for "Containerutleie - Stort utvalg!" (Container rental - Large selection!) featuring various storage containers. To the right of the article, there is a sidebar titled "Les videre" with two more news items:

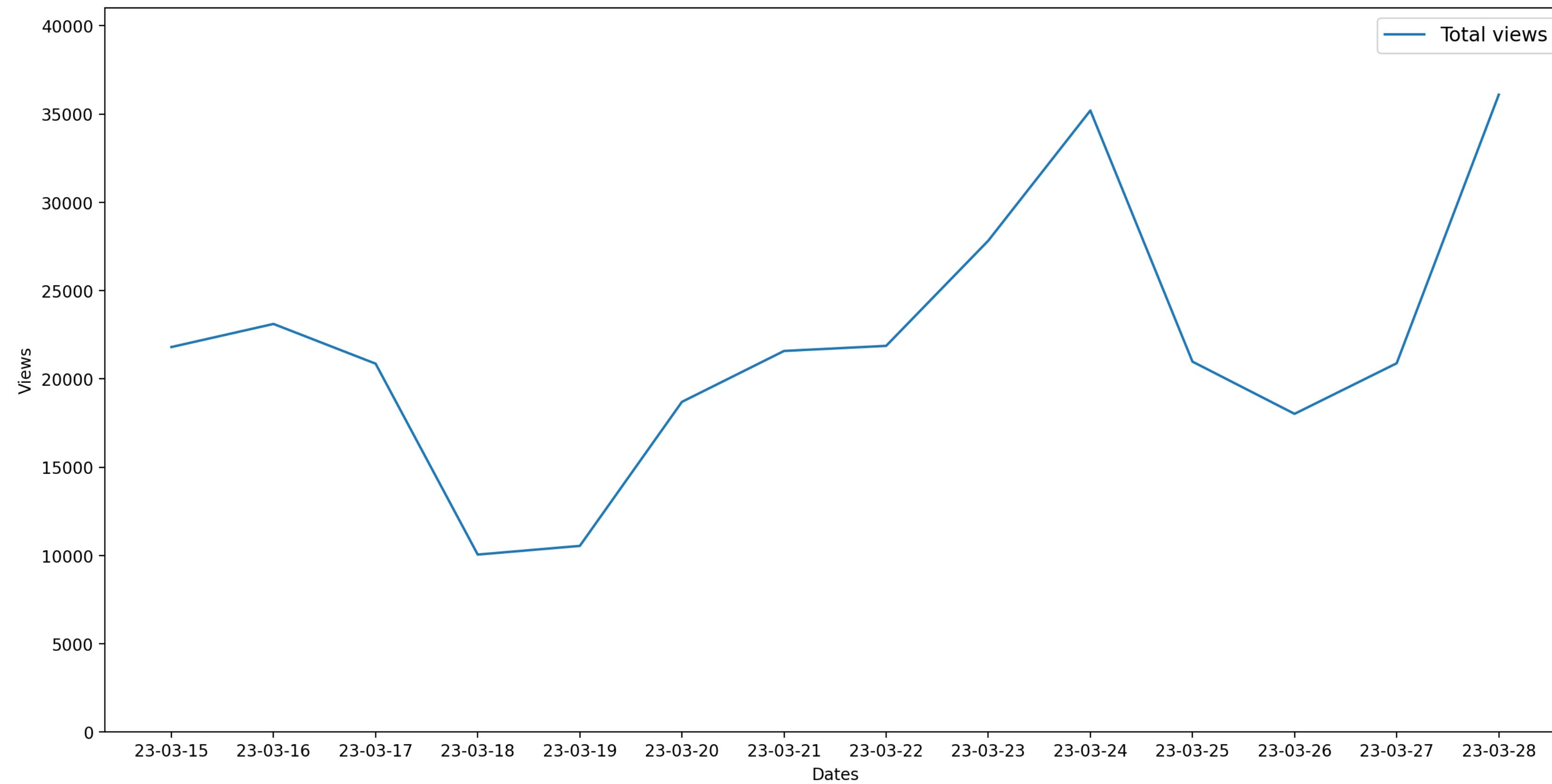
- «Brann-kjøpet er 24 år. Han er altså for gammel.»** (Brann-kjøpet is 24 years old. He is therefore too old.) - Date: 09. MARS
- Finne satt på benken i stillhet: – Betyr ikke at jeg ikke blir forbannet** (Finne sat on the bench in silence: – It doesn't mean I'm not banned.) - Date: 19. FEBRUAR

At the bottom of the page, there are two images of football players in red uniforms, one from each side of a match.

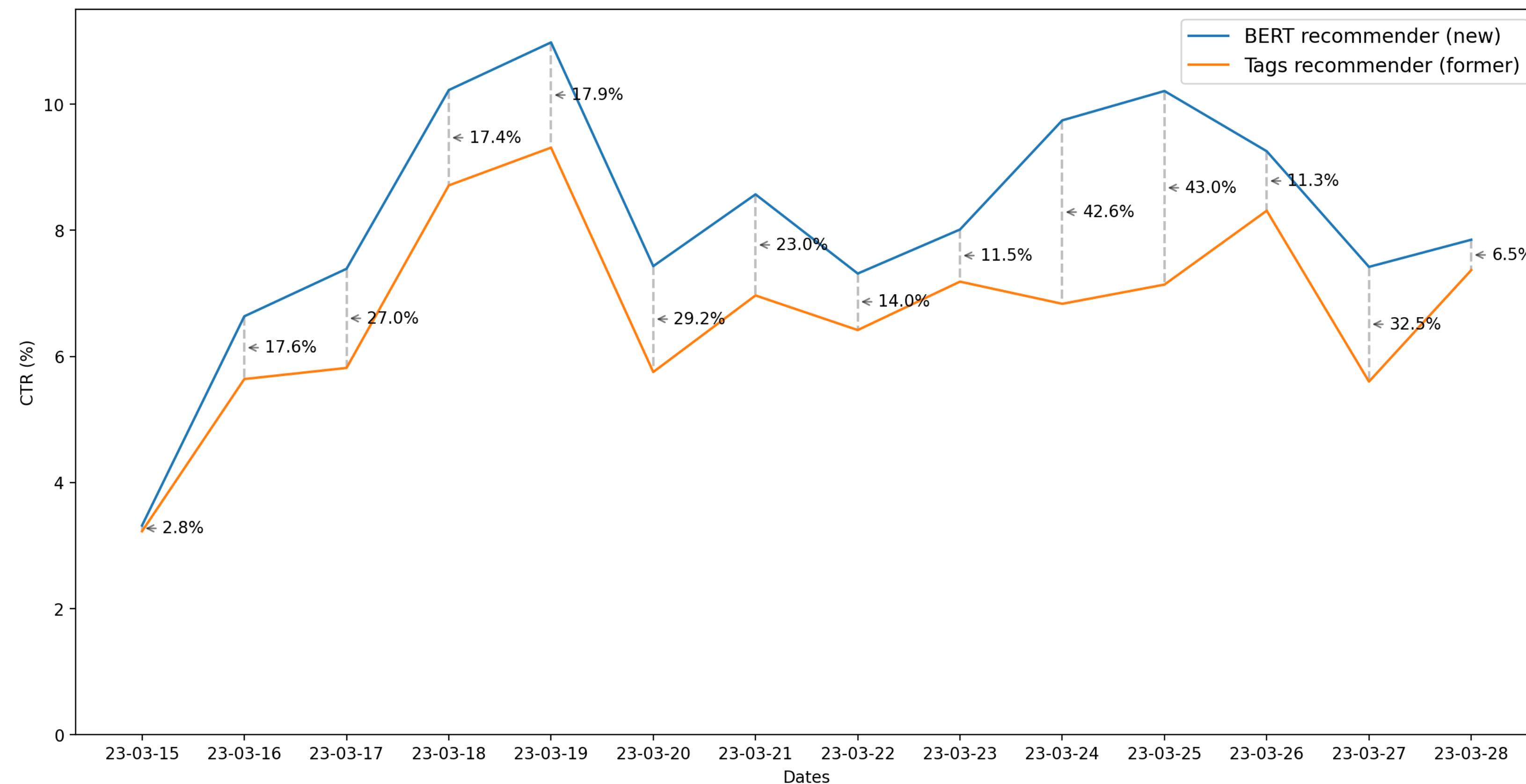
Clicks performed on the A/B test



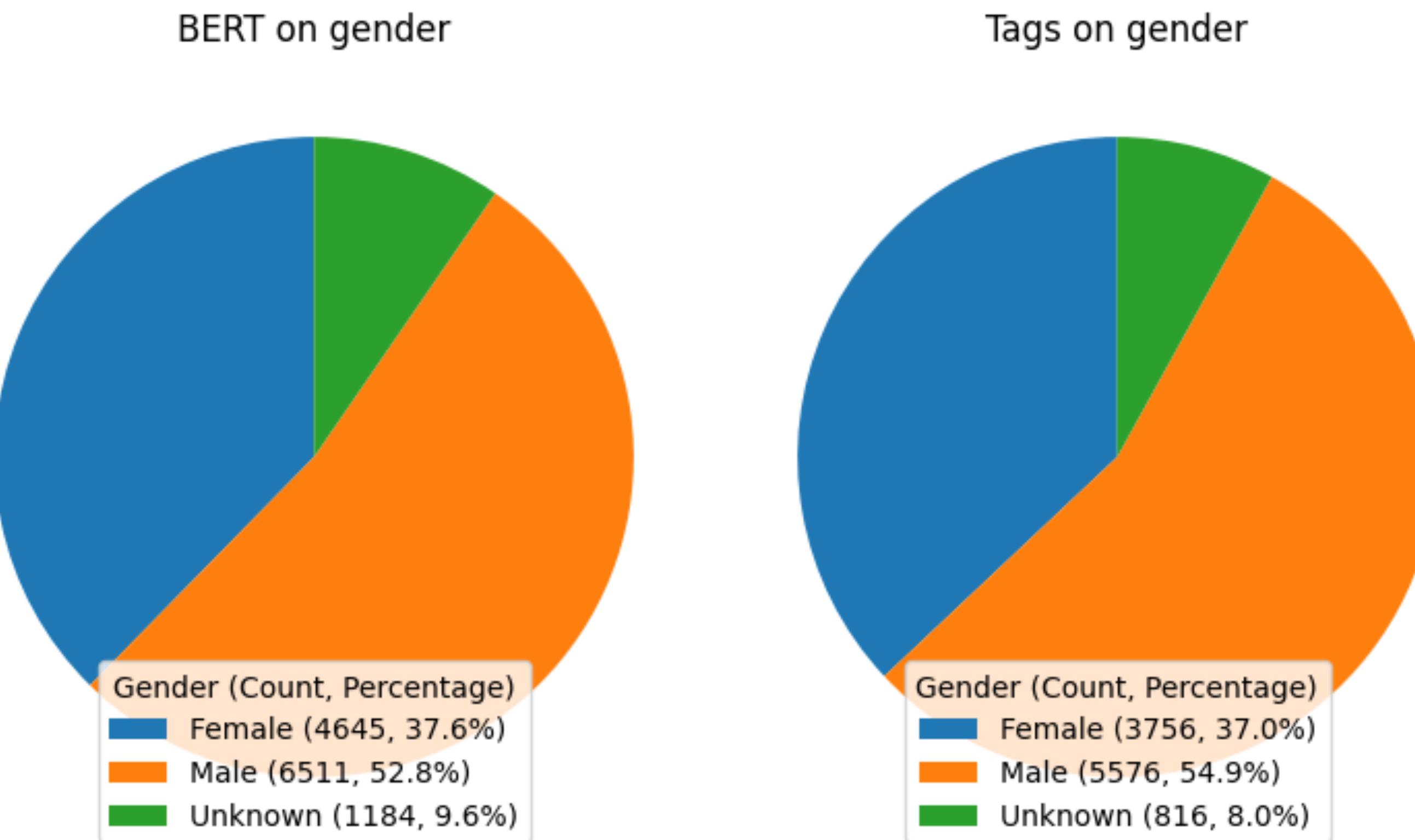
Views on the A/B test



CTR on the A/B test

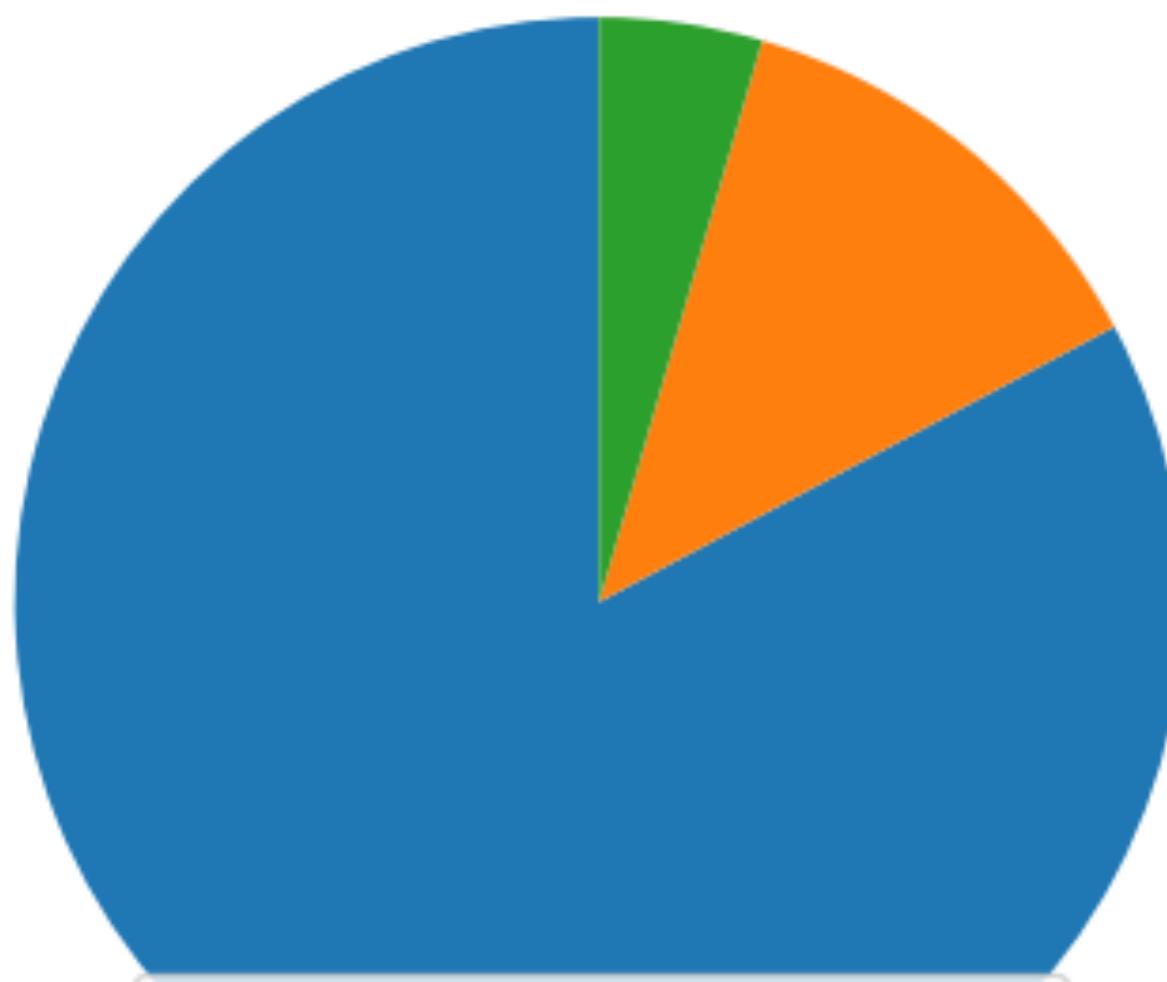


Gender distribution of user clicks of A/B test



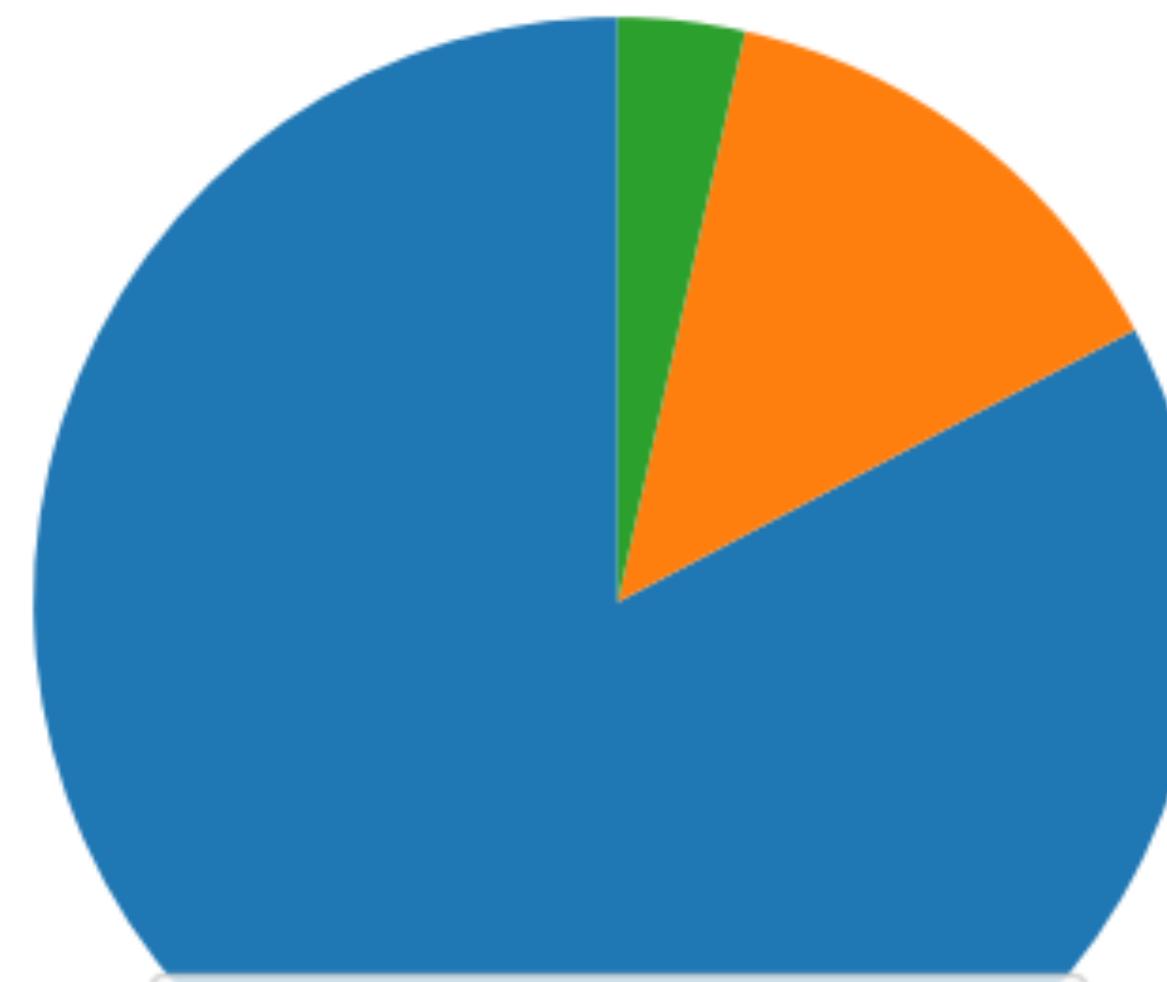
Age distribution of user clicks of A/B test

BERT on age



| Age (Count, Percentage) | |
|-------------------------|----------------|
| Age > 45 | (10218, 82.8%) |
| Age <= 45 | (1566, 12.7%) |
| Age = Unknown | (556, 4.5%) |

Tags on age



| Age (Count, Percentage) | |
|-------------------------|---------------|
| Age > 45 | (8393, 82.7%) |
| Age <= 45 | (1400, 13.8%) |
| Age = Unknown | (355, 3.5%) |

Findings of A/B test

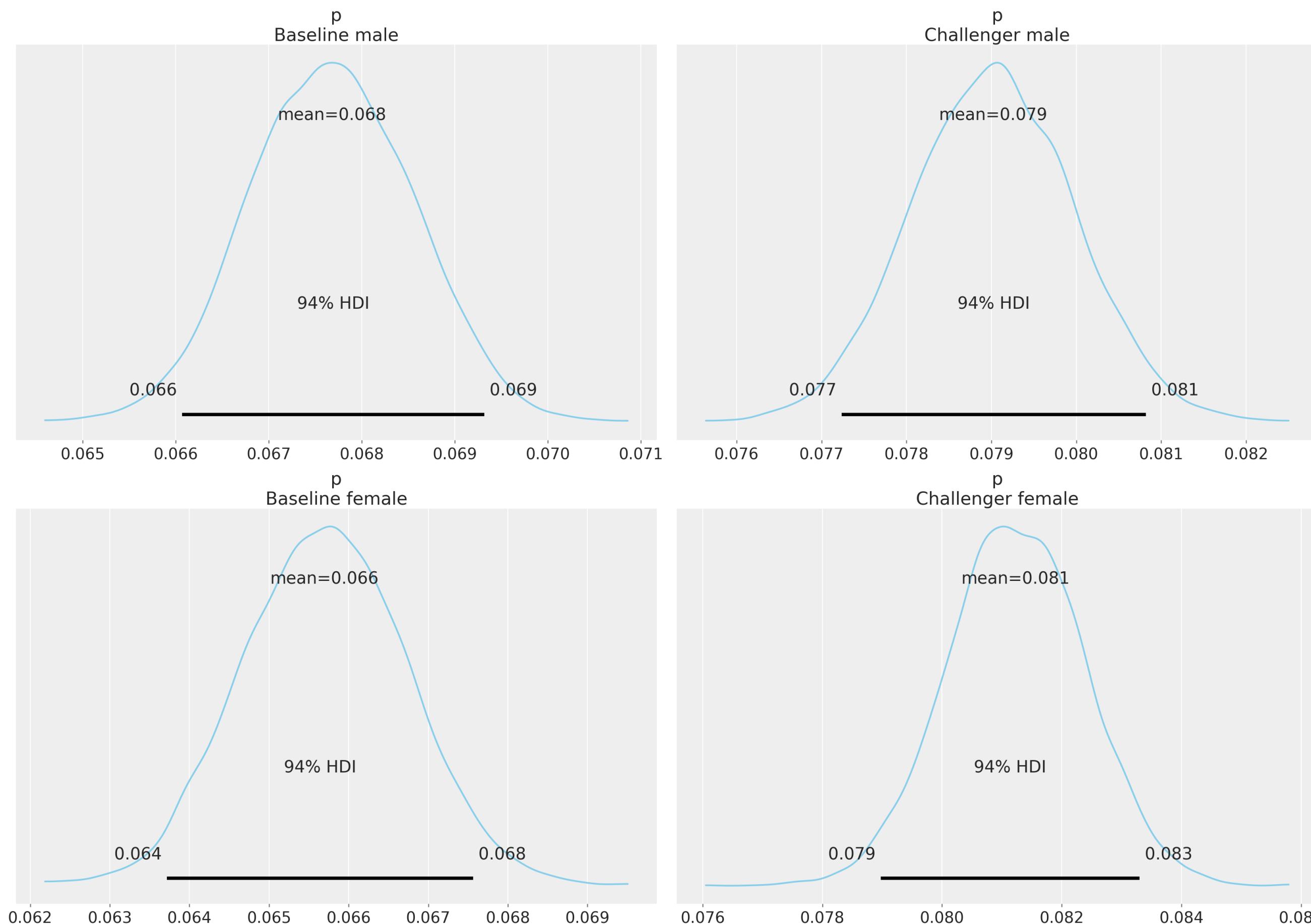
- **11 528** participants
 - BERT – 5 688
 - Tags – 5 144
 - Both - 696
- **22 488** total clicks
 - BERT 12 340
 - Tags 10 148
- BERT recommendations **21.6%** greater engagement
- Male and age > 45

Bayesian statistics

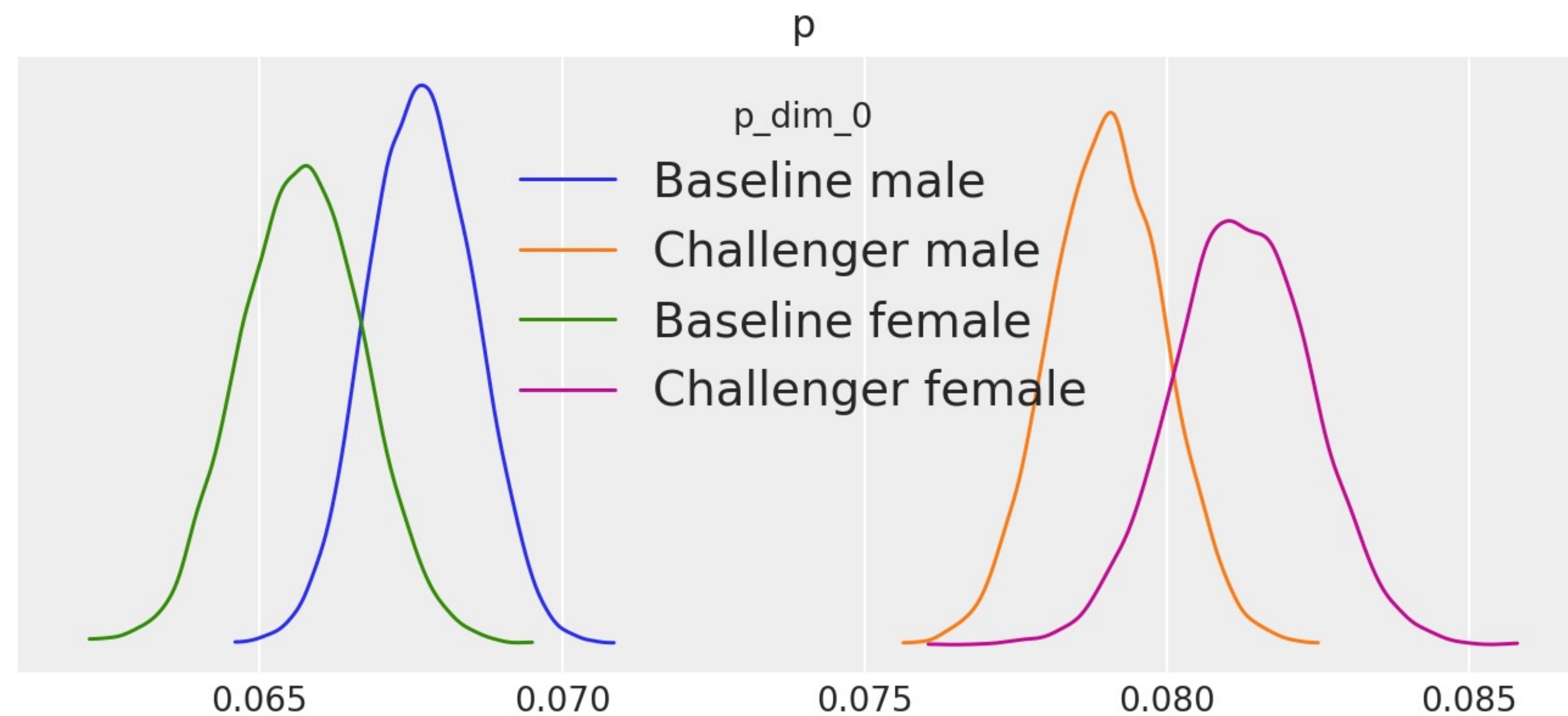
- Principle of probability statements and updating the probabilities after collecting new data
- Binomial distribution
 - Trials (views)
 - Successes (clicks)
- $P(\lambda)$ = prior distribution
 - Beta(1,1) – flat prior
- $P(X | \lambda)$ = likelihood of data X given parameter λ
- $P(X)$ = marginal likelihood of data X

$$P(\lambda|X) = \frac{P(\lambda)P(X|\lambda)}{P(X)}$$

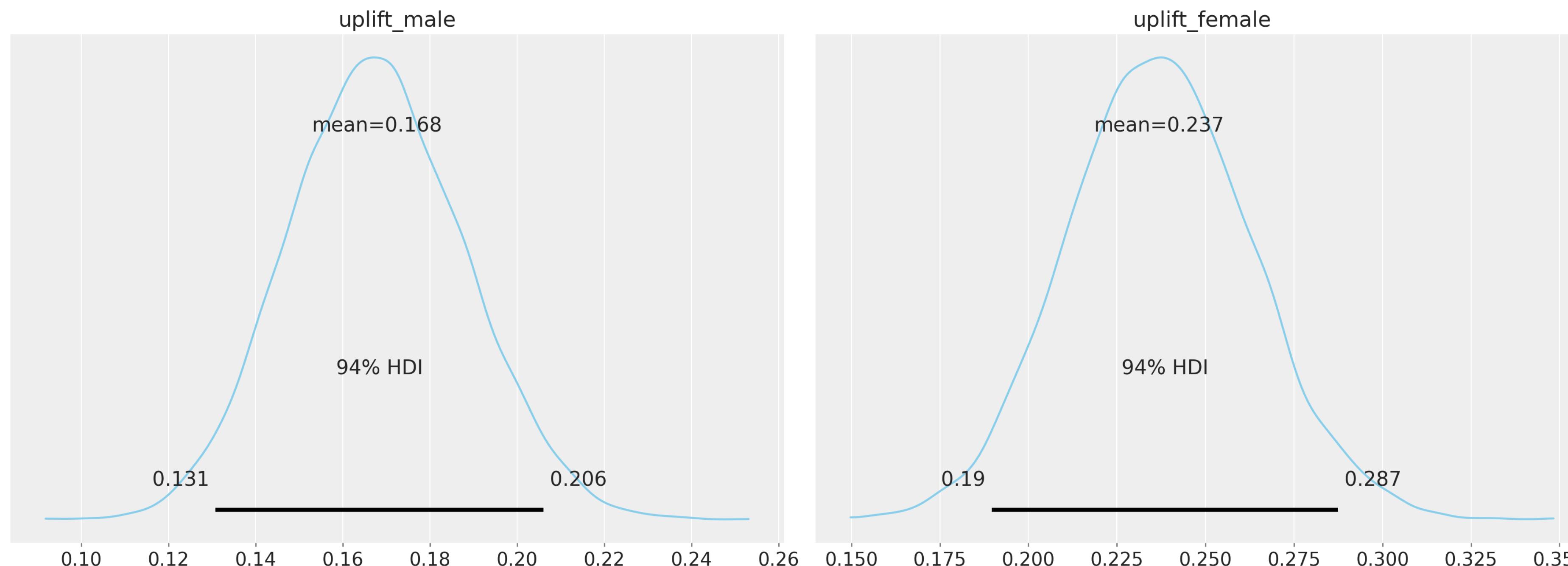
Gender posterior distributions



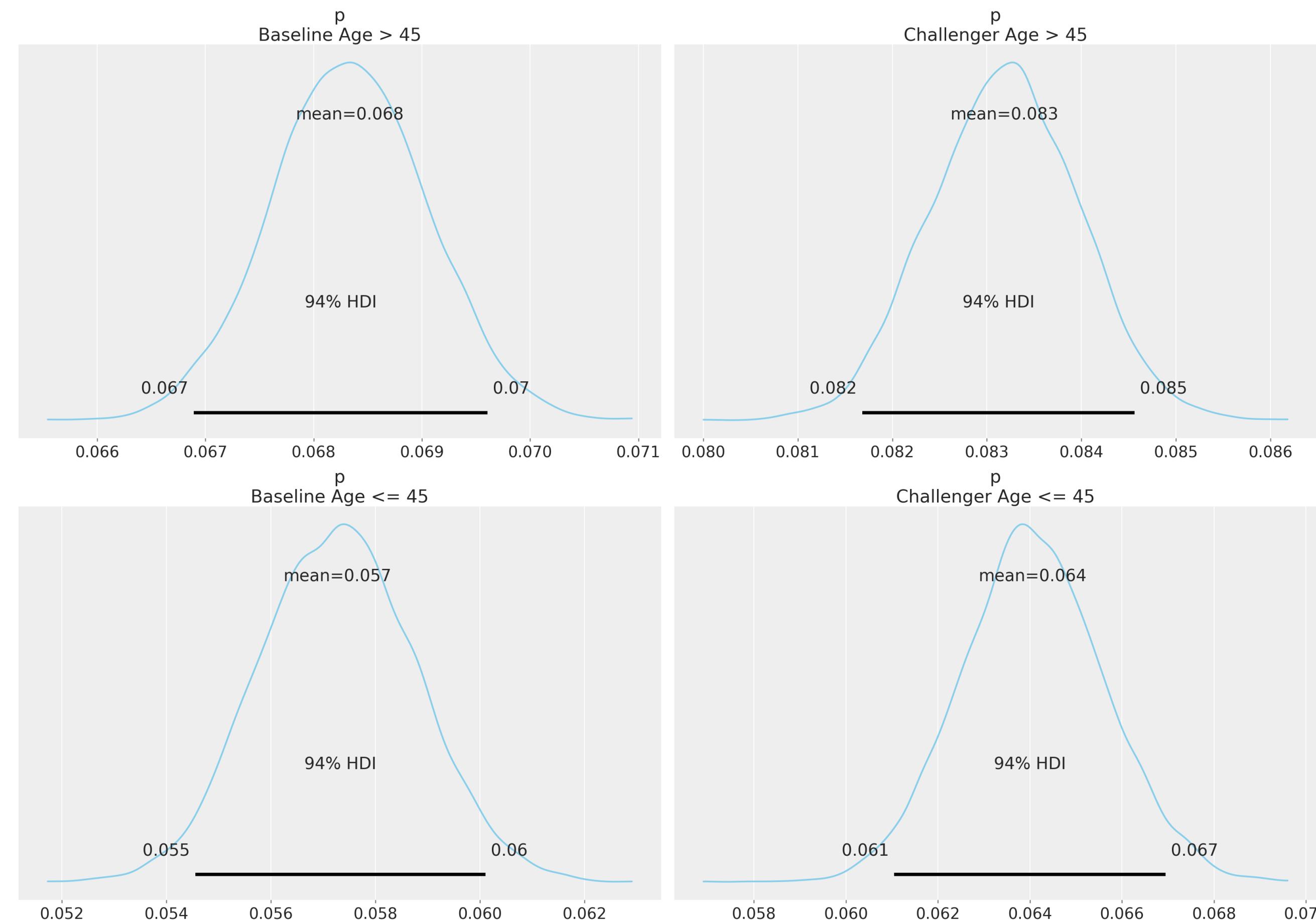
Summary of gender posterior distributions



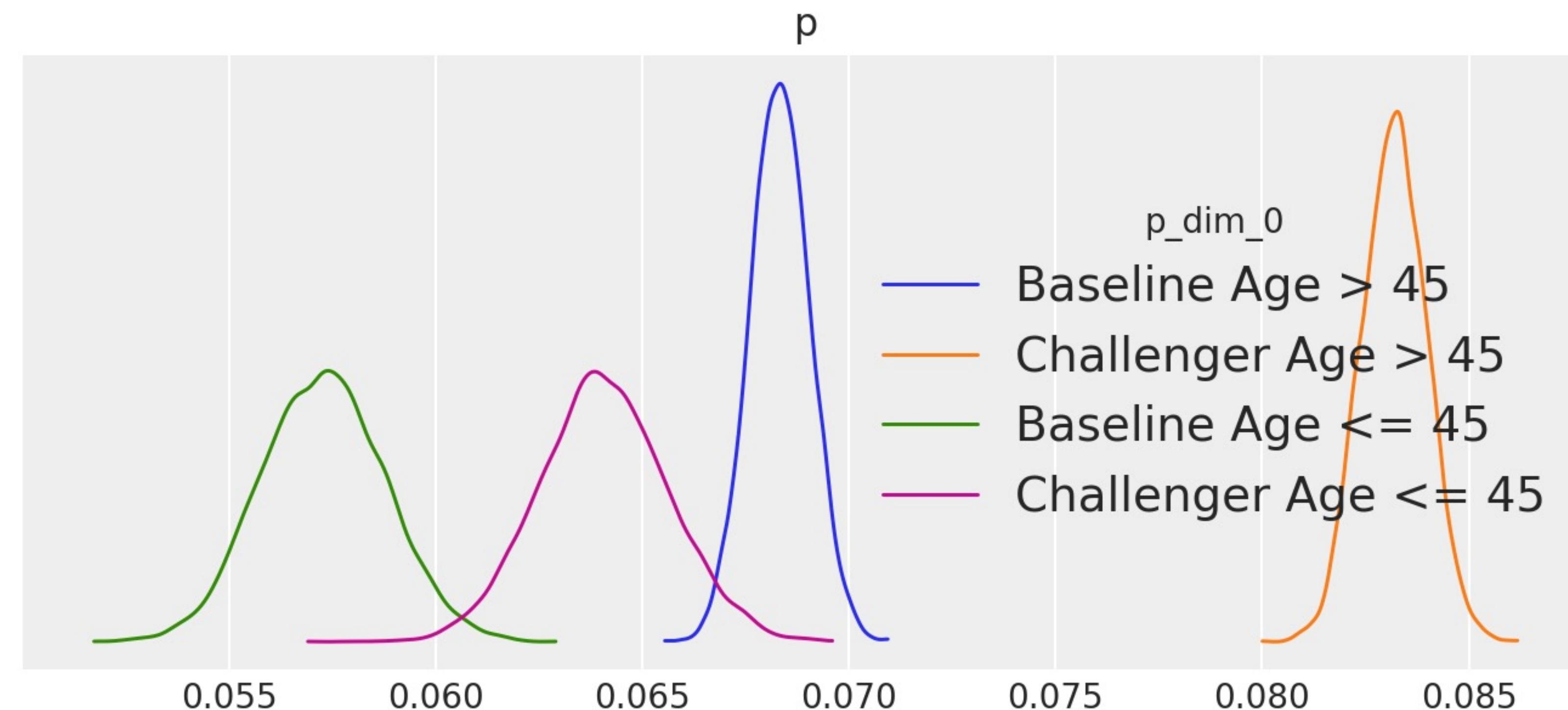
Uplift gender posterior distributions



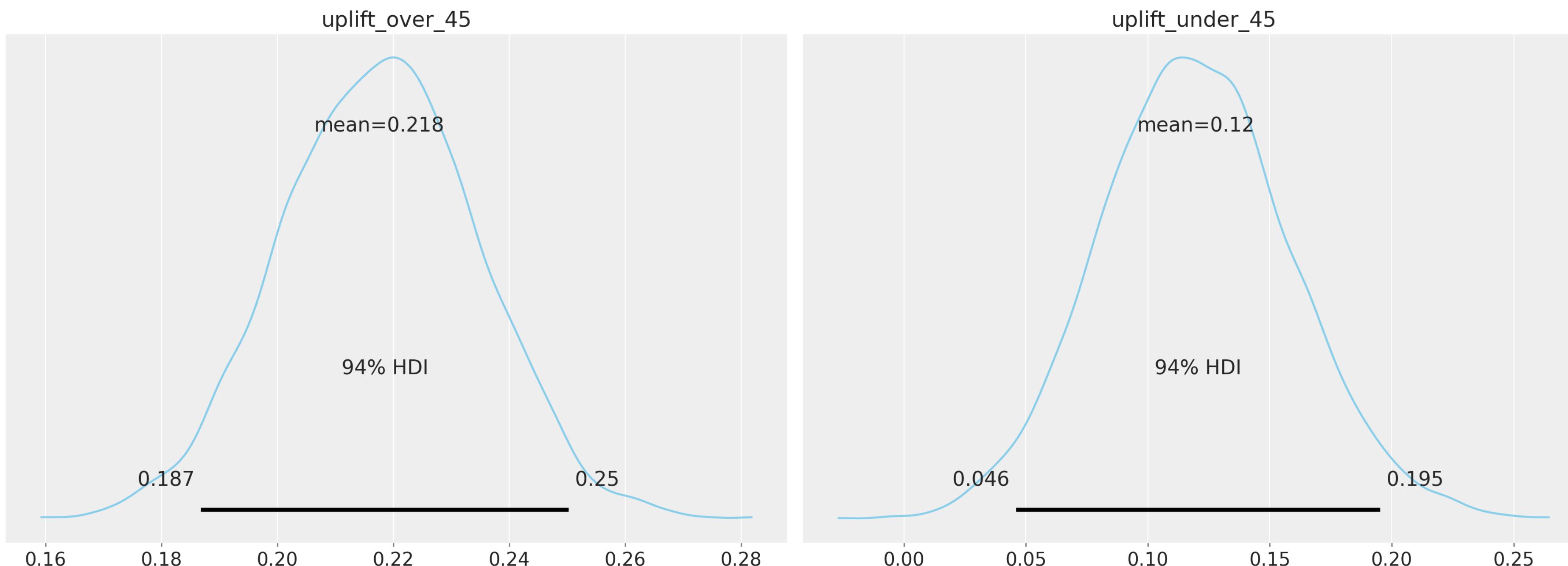
Age posterior distributions



Summary of age posterior distributions



Uplift age posterior distributions



Results and why an Bayesian approach is chosen

- BERT(Challenger) is better in **almost all** cases
- More **relatable** for non-technicians
- Traditional statistics may have too much **statistical power**
- Exact **p-value** is not stated
 - Rather posterior distributions
 - 94% credible intervals

Summary

- Increased performance of ML models
- The importance of filters and hyperparameters
- TSNE dimension reduction
 - BERTs contextual embedding
 - Ability to cluster quite well
- Greater engagement of BERT recommendations

Futura Media

Thank you for your attention

Contact information:

Peter Kolbeinsen Klingenberg

Peter.klingenbergs@student.uib.no

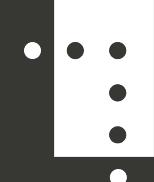
Research Centre for Responsible Media

Technology and Innovation

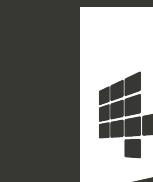
Project number 309339



Norwegian Centre
for Research-based
Innovation



Nasjonalbiblioteket



amedia Bergens Tidende HIGHSOFT