

# Explaining rankings

Maartje ter Hoeve  
University of Amsterdam & Blendle

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
Conclusion

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
conclusion

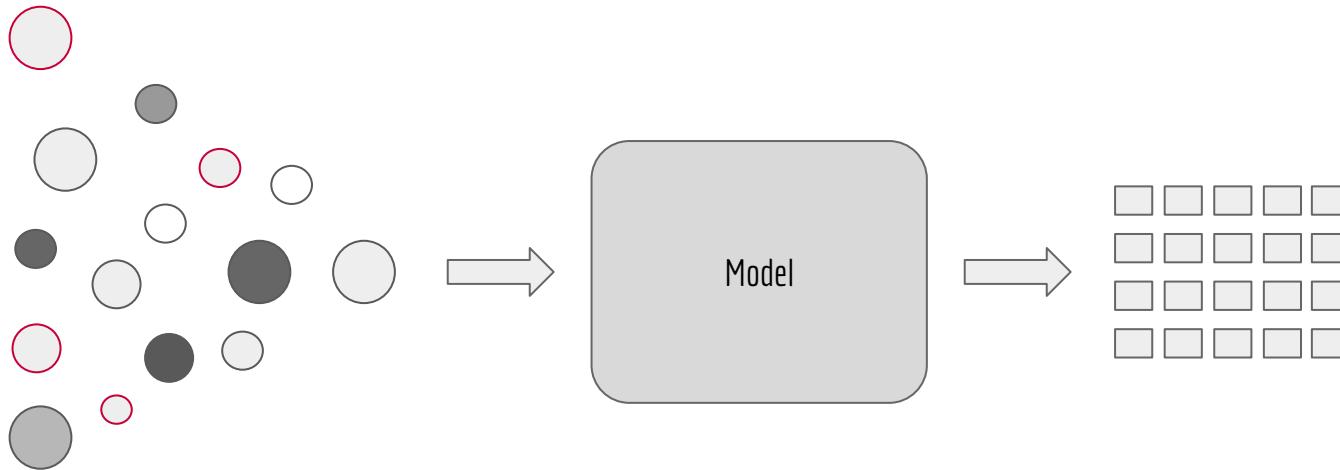
# Introduction

---

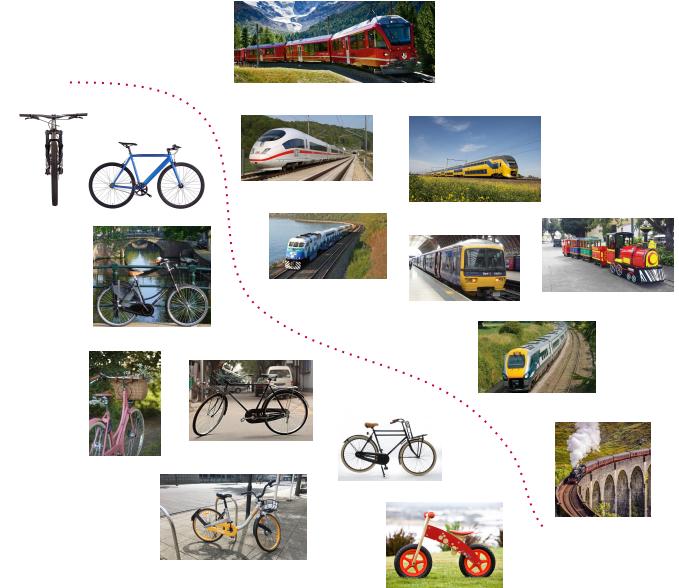
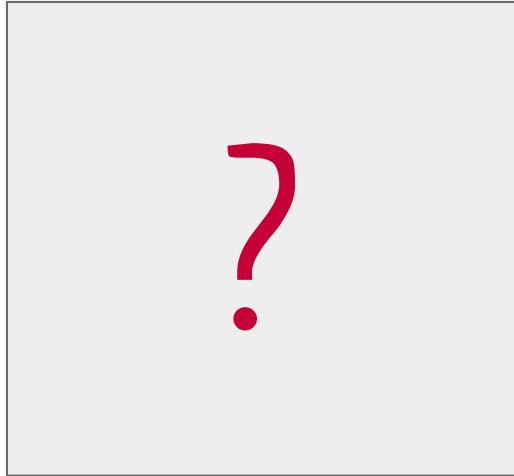
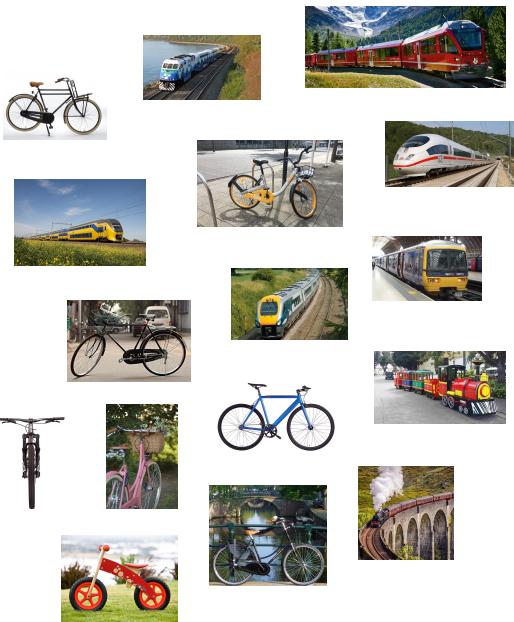
**What** is explainability and **why** is it needed?

# Explainability: what and why?

---



# Explainability: what and why?



# Explainability: what and why?

---



# Explainability: what and why?

---



Maartje ter Hoeve

✉ maartje.terhoeve@student.uva.nl



# But what is an explanation?

---

An explanation needs to **faithfully** give the underlying **cause** of an event

# But what is an explanation?

---

## Justification

Provide conceptual explanations that do not necessarily expose the underlying structure of the algorithm

## Description

Provide conceptual explanations that do expose the underlying structure of the algorithm

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
conclusion

# Rankings

- 1
- 2
- 3
- 4
- ⋮
- n - 2
- n - 1
- n



Afbeeldingen van ranking



→ Meer afbeeldingen voor ranking



Afbeeldingen melden

# How do we explain a ranking?

---

231	543
228	432
203	398
157	231
⋮	⋮
6	8
3	4
1	2

Only looking at the score of an item is not sufficient

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
conclusion

# Blendle

## Maandag

Zo krijg je **meer plezier** in je werk, hulpverleners Bangladesh **verbijsterd** door 'mensenplaag', Khadija Arib probeert te verklaren **waar het misging** tussen de Nederlanden en de nieuwkomer én een **portret** van Energiecommissaris Ruud Koomstra.



Zo krijg je meer plezier in je werk (en je collega's ook)



Voor jou



# Blendle



Blendle already has heuristic **justifications**

We use these as one of our baselines

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
conclusion

# Research questions

---

## PART 1

RQ1 Do users want to receive explanations of why particular news items are recommended to them?

RQ2 What way of showing news recommendations reasons do users prefer: textual or visual reasons; a single reason or multiple reasons; apparent or less apparent reasons?

# Research questions

---

## PART 2

**RQ3** How do we provide users with easy to understand, uncluttered, listwise explanations?

**RQ4** How do we build an explanation system that produces faithful, model-agnostic explanations for the outcome of a ranking algorithm, yet is scalable so that it can run in real time?

**RQ5** Does the reading behaviour of users who are provided with model-agnostic listwise explanations for a personalized ranked selection of news articles differ from the reading behaviour of users who are provided with heuristic or pointwise explanations for a personalized ranked selection of news articles?

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

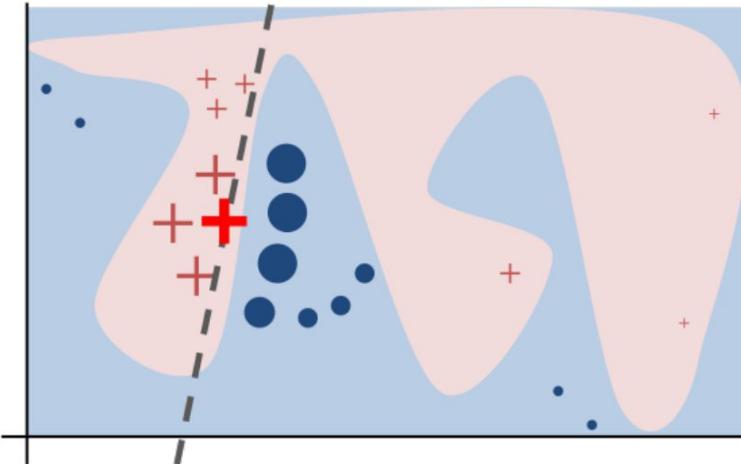
Answering  
research  
questions

Discussion and  
conclusion

# Related work

---

LIME (Ribeiro et al, 2016): find a local, faithful explanation for the decision of any classifier



LIME is a baseline of this research

We work with rankings, not classifiers

Therefore we bin our ranking scores

mLIME

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
conclusion

# Part 1 - User study



Single Reason - Visible



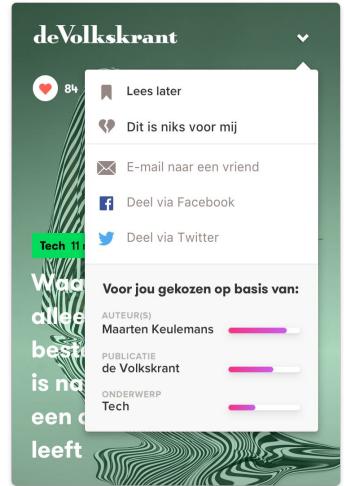
Single Reason - Invisible



Multiple Reasons - Visible



Multiple Reasons - Combined



Bar chart

# RQ1 Do users want explanations?

- 1 → Op het plaatje hieronder zie je hoe een artikel op Blendle Premium er nu uitziet. De artikelen die je ziet worden voor je uitgekozen op basis van je persoonlijke voorkeuren en wat je graag leest.

Stel, we zouden je meer informatie geven over waarom we een artikel voor je hebben uitgezocht. Zou je dat nuttig vinden?\*



Ja     Een beetje     Nee  
 Weet ik niet

User wants reasons	Times answered
Yes	65
Somewhat	24
No	26
I don't know	5

$$\chi^2 = 14.55, p < 0.001$$

# RQ2 Preferences how explanations are shown?



Single Reason - Visible



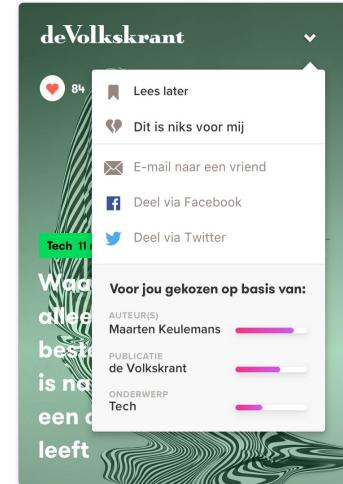
Single Reason - Invisible



Multiple Reasons - Visible



Multiple Reasons - Combined

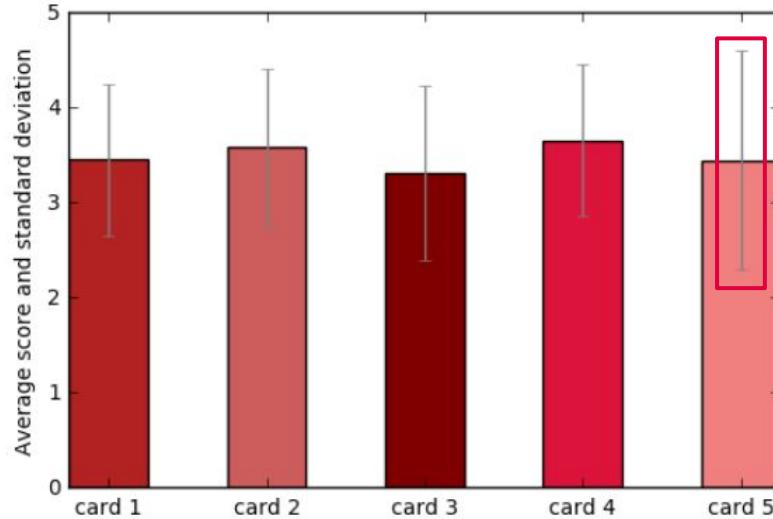


Bar chart

# RQ2 Preferences how explanations are shown?

Transparency  
Sufficiency  
Trust  
Satisfaction

}

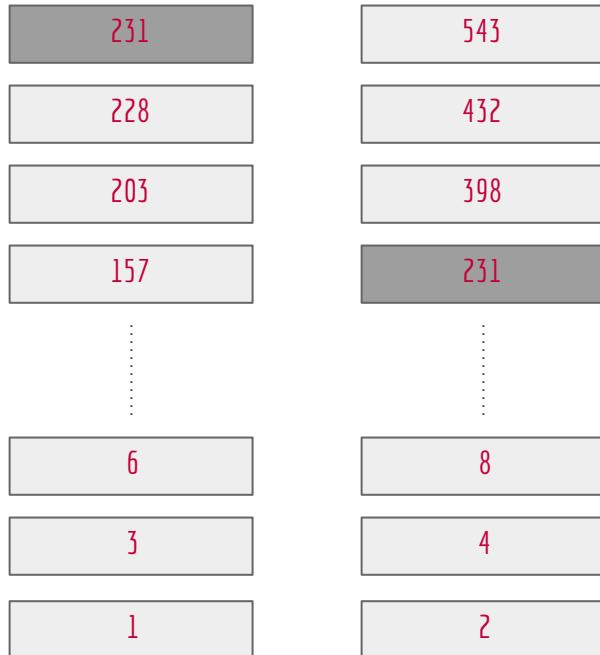


# Part 2

---

# RQ3 How do we make listwise explanations?

---



Explain the entire list?

Explain items in comparison to other items?

Explain which features were important for the position in the ranking?

# RQ3 How do we make listwise explanations?

---

231	543
228	432
203	398
157	231
⋮	⋮
6	8
3	4
1	2

Explain the entire list?

Explain items in comparison to other items?

Explain which features were important for the position in the ranking?

# RQ3 How do we make listwise explanations?

---

Which features are most important for the item's position in the ranking?

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

⋮

⋮

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

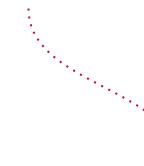
f0	f1	f2	f3	f4
----	----	----	----	----

# RQ4 How do we design this?

---

## Main intuition

If we change feature values and the ranking changes, then this feature was important



$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

---

## Training phase

Find how feature values change the ranking

Find disruptive distributions and points of interests

## Explaining phase

Use distributions to sample feature values from

Find most important features

Return most important features as explanations

LISTEN - LISTwise ExplaiNer

# RQ4 How do we design this?

---

Training phase

# RQ4 How do we design this?

---

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

# RQ4 How do we design this?

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

1	$f_1$	$f_2$	$f_3$	$f_4$	90
---	-------	-------	-------	-------	----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

200

180

90

120

100

$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

200

180

160

120

100

$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

4	$f_1$	$f_2$	$f_3$	$f_4$	170
---	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

5	$f_1$	$f_2$	$f_3$	$f_4$	160
---	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

200

180

160

120

100

190

$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

Find how feature values change the ranking:  $f_0 [1, 2, 3, 4, 5, 6]$

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	200
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	180
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	160
-------	-------	-------	-------	-------	-----

⋮

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	120
-------	-------	-------	-------	-------	-----

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	100
-------	-------	-------	-------	-------	-----

200

180

160

120

100

200

180

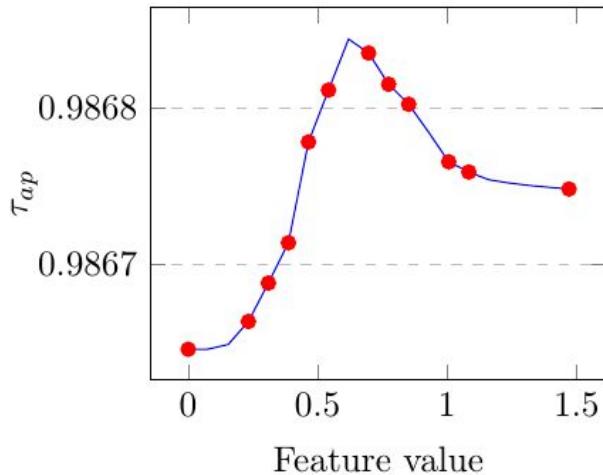
195

$$\tau_{AP} = \frac{2}{N-1} \sum_{i=2}^N \left( \frac{C(i)}{i-1} \right) - 1$$

# RQ4 How do we design this?

---

Find disruptive distributions and points of interests



# RQ4 How do we design this?

---

Explanation phase

# RQ4 How do we design this?

Find most important features

f0	f1	f2	f3	f4	200
----	----	----	----	----	-----

f0	f1	f2	f3	f4	180
----	----	----	----	----	-----

f0	f1	f2	f3	f4	160
----	----	----	----	----	-----

⋮

f0	f1	f2	f3	f4	120
----	----	----	----	----	-----

f0	f1	f2	f3	f4	100
----	----	----	----	----	-----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

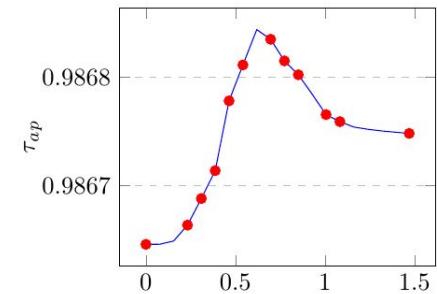
170

180

160

120

100



# RQ4 How do we design this?

---

Return most important features

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

⋮

f0	f1	f2	f3	f4
----	----	----	----	----

f0	f1	f2	f3	f4
----	----	----	----	----

# RQ4 How do we design this?

---

Is this faithful?

Construct dummy data where we know what to expect

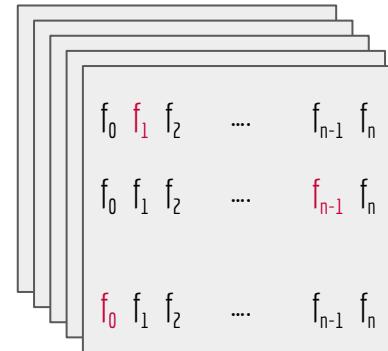
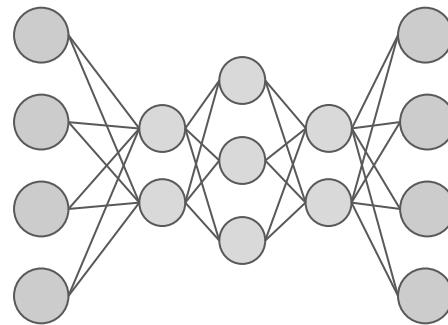
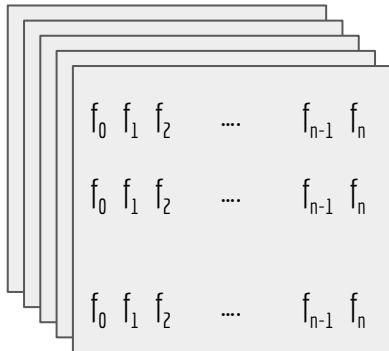
LISTEN gives the correct results

Disruptive distribution approach slightly decreases faithfulness, yet increases speed

# RQ4 How do we design this?

This is still not fast enough to run in production

Q-LISTEN

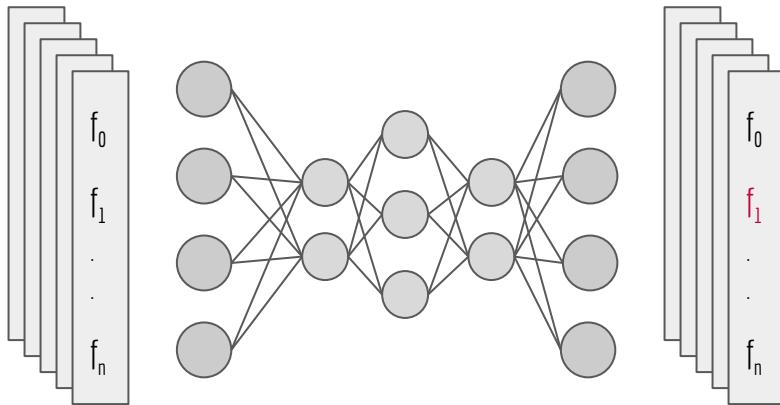


# RQ4 How do we design this?

---

Same holds for mLIME

Q-mLIME



# RQ5 How do users behave?

---

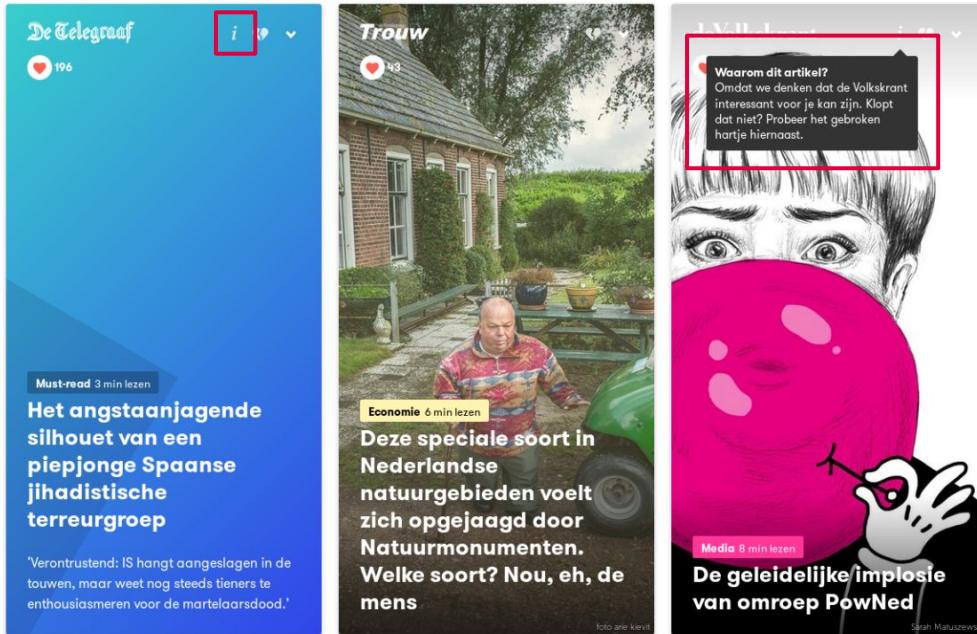
A/B/C - test

A. Heuristic

B. Q-mLIME

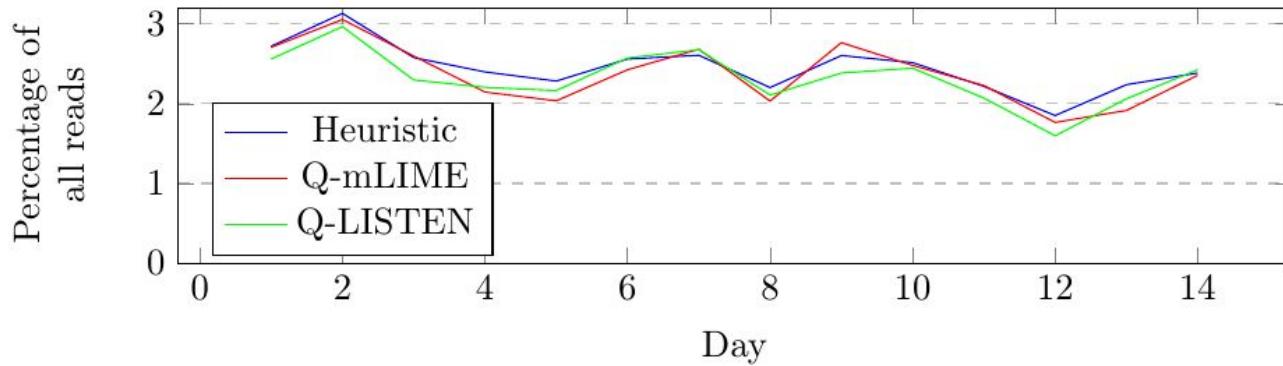
C. Q-LISTEN

# RQ5 How do users behave?



# RQ5 How do users behave?

## Results



# RQ5 How do users behave?

---

## Results

	Heuristic	Q-mLIME	Q-LISTEN
% reasons seen (of total reads)	3.55	3.54	3.51
% reasons seen (of total reasons)	34.1	32.9	32.0
Reasons per user (of users that see reasons)	1.83	1.86	1.72
Article opened within 2 minutes (% of all reasons seen)	12.1	12.9	10.6
Reasons seen after 20 minutes of opening (% of all reasons seen)	11.0	10.1	10.7

# Content

---

What and why?

Rankings

Blendle

Research  
questions

Related work

Answering  
research  
questions

Discussion and  
conclusion

# Conclusion and discussion

---

## PART 1

**RQ1** Do users want to receive explanations of why particular news items are recommended to them?

Yes

**RQ2** What way of showing news recommendations reasons do users prefer: textual or visual reasons; a single reason or multiple reasons; apparent or less apparent reasons?

No specific one

# Conclusion and discussion

---

## PART 2

RQ3 How do we provide users with easy to understand, uncluttered, listwise explanations?

We look at which features were most important for an item's position in the ranking

RQ4 How do we build an explanation system that produces faithful, model-agnostic explanations for the outcome of a ranking algorithm, yet is scalable so that it can run in real time?

We find feature importance by changing the feature values and changing the rankings

We make disruptive distributions

We learn the explanation space with a neural net

# Conclusion and discussion

---

**RQ5** Does the reading behaviour of users who are provided with model-agnostic listwise explanations for a personalized ranked selection of news articles differ from the reading behaviour of users who are provided with heuristic or pointwise explanations for a personalized ranked selection of news articles?

We do not find any significant differences

# Conclusion and discussion

---

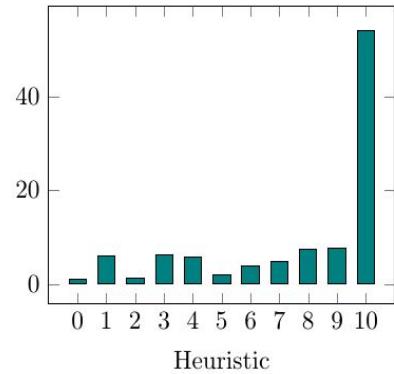
## Take home message

Users clearly state that they want explanations. Therefore, even though their behaviour is not affected by the explanations, still provide them with faithful explanations

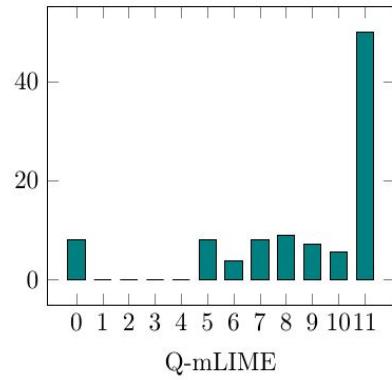
# Reasons seen

---

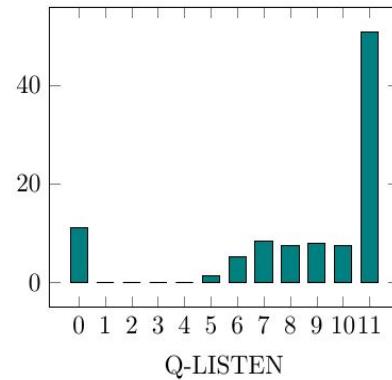
Percentage reason was clicked



Percentage reason was clicked

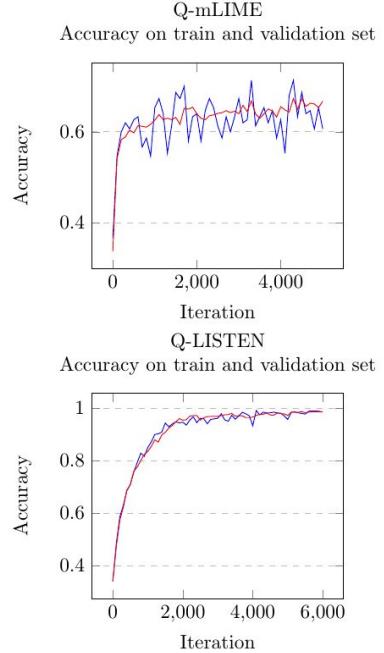
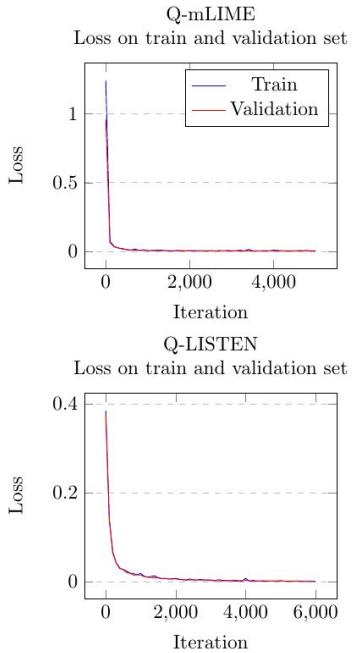


Percentage reason was clicked



# Loss and accuracy curves

---



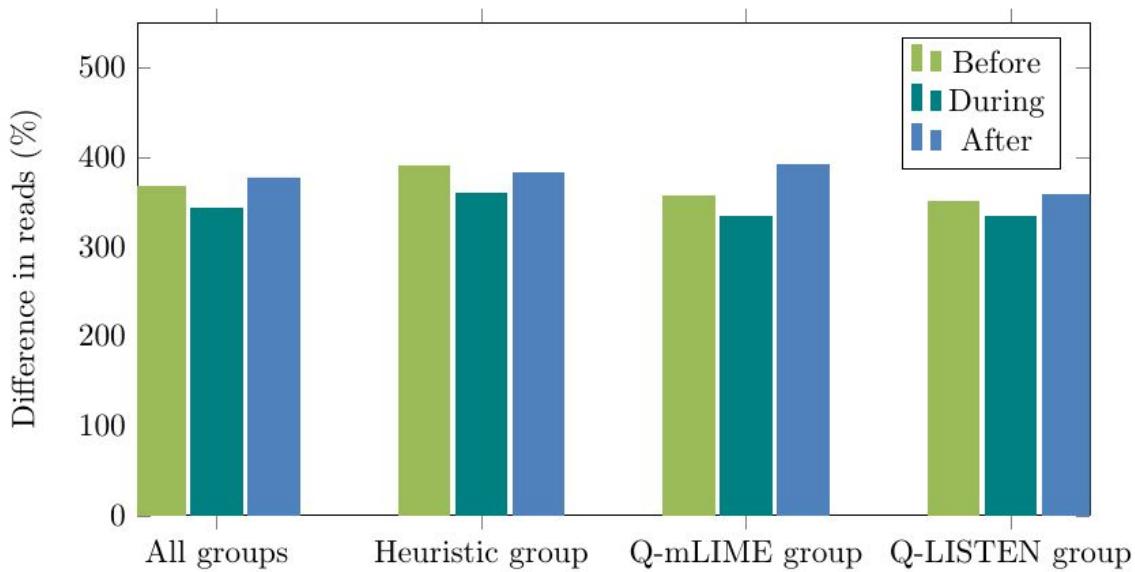
# mLIME vs LISTEN

---

Position item in ranking	mLIME or LISTEN	$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$
1	mLIME	█			█					
	LISTEN	█			█				█	
2	mLIME	█		█	█					
	LISTEN	█			█				█	
15	mLIME		█							
	LISTEN	█						█		

# Active and less active users

---



# RQ5 How do users behave?

## Donderdag

Zo veel vriendschappen kun je maximaal onderhouden, PowNed-baas verdedigt zijn controversiële omroep en waarom we graag nepnieuws delen.

