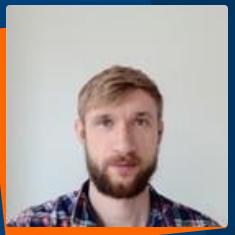




idealo's Journey in Building a Production Learning to Rank System

MICES 2025



Gennady Shabanov



Machine Learning Engineer

Atakan Filgöz

Machine Learning Engineer

Agenda

- Introducing idealo
- Problem Statement
- The Solution
- The Architecture
- Challenges and Potential Solutions
- Stats

About idealo



Germany's 4th largest
ecommerce website



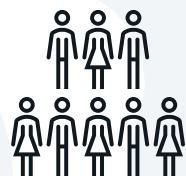
50.000+ shops



Active in 6 countries
(DE, AT, ES, IT, FR, UK)



5M+ products
500M+ offers



18M+ app downloads



80M+ visits per month

Product Catalog

idealo

≡ **25** **idealo** JAHRE

iphone 16 pro max

Merkzettel | Preiswecker | Anmelden

Home > Elektroartikel > Telekommunikation > Handys & Smartphones

Handys & Smartphones (441)* ⓘ

Sortieren nach: Beliebteste zuerst

Ergebnisse in Handys & Smartphones werden angezeigt. Ergebnisse in allen Kategorien anzeigen.

Produktfinder ⓘ

Welche Marke bevorzugst du für dein neues Smartphone?

Samsung Apple Nicht wichtig

Samsung und Apple sind die beliebtesten Marken, bieten aber unterschiedliche Betriebssysteme und Funktionen.

Preis Zurücksetzen

1 € - 2610 €

bis 510 € 110
 530 € bis 790 € 142
 790 € bis 1.400 € 129
 ab 1.400 € 60

Verfügbarkeit

Schnell lieferbar (1-7 Werkstage) 329

Hersteller

Apple iPhone 16 Pro Max 256GB Titan Schwarz

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, iOS 18, 4.685 mAh

Note ⚭ 1,7 ★★★★★ 10

916 Angebote

ab 1.109,90 €

Produktdetails

Apple iPhone 16 Pro Max 256GB Titan Wüstensand

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, 256 GB interner Speicher, 512 GB interner Speicher

★★★★★ 1

90 Angebote

ab 1.111,70 €

Produktdetails

Apple iPhone 16 Pro Max 256GB Titan Natur

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, 256 GB interner Speicher, 512 GB interner Speicher

★★★★★ 3

95 Angebote

ab 1.109,90 €

Produktdetails

Apple iPhone 16 Pro Max 256GB Titan Weiß

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, 256 GB interner Speicher, 512 GB interner Speicher

88 Angebote

ab 1.113,90 €

Produktdetails

Apple iPhone 16 Pro Max 512GB Titan Schwarz

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, 512 GB interner Speicher, 512 GB interner Speicher

70 Angebote

ab 1.327,90 €

Produktdetails

Apple iPhone 16 Pro Max 512GB Titan Wüstensand

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, 512 GB interner Speicher, 512 GB interner Speicher

71 Angebote

ab 1.339,00 €

Produktdetails

Apple iPhone 16 Pro Max 1TB Titan Schwarz

5G Handy, 6,9 Zoll, Full HD, 48 MP, 8 GB RAM, 1.000 GB interner Speicher, 1.000 GB interner Speicher

116 Angebote

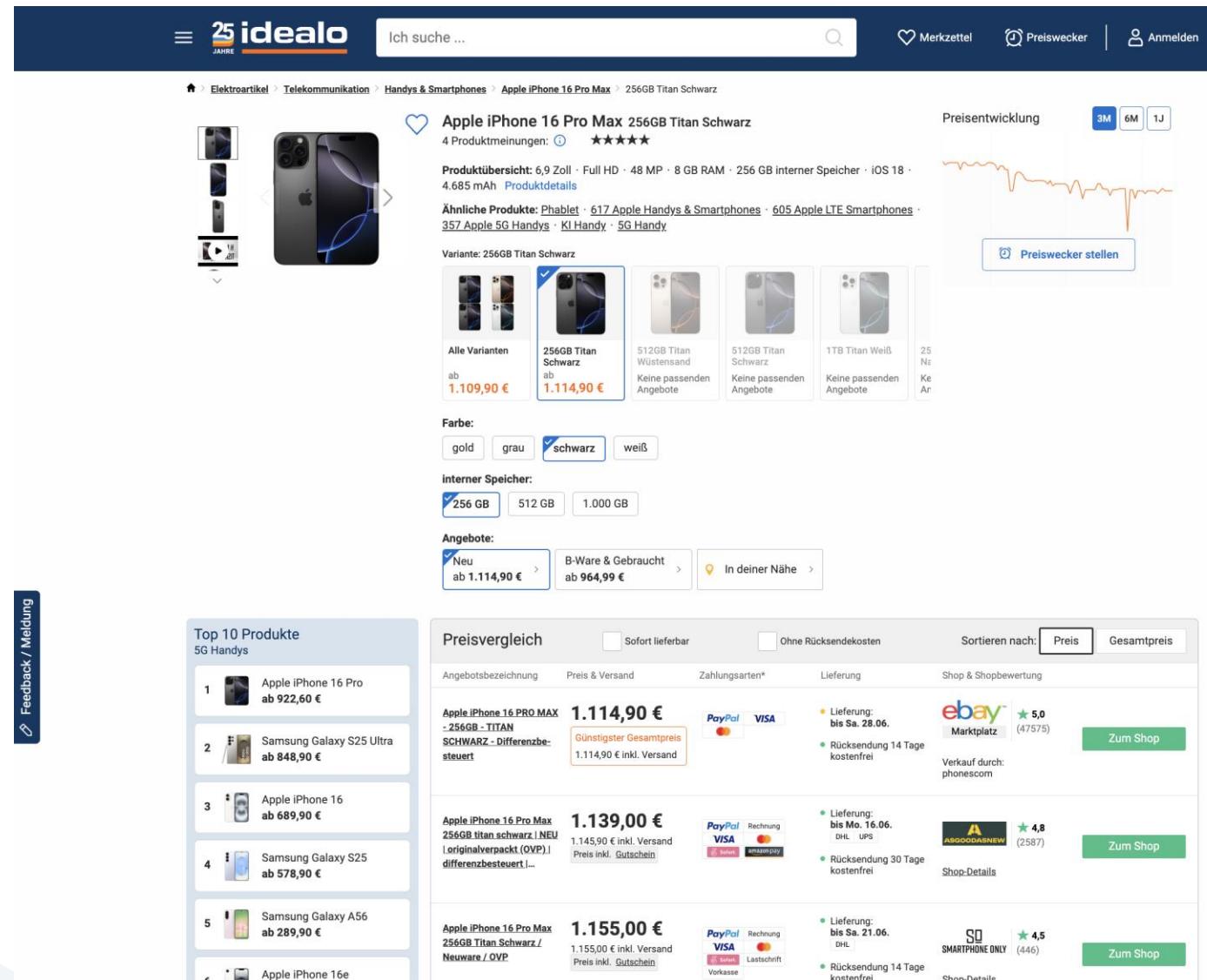
ab 1.495,00 €

Produktdetails

Feedback / Meldung

Product Catalog - Offers

idealo



25 Jahre idealo

Ich suche ...

Merkzettel

Preiswecker

Anmelden

Elektroartikel > Telekommunikation > Handys & Smartphones > Apple iPhone 16 Pro Max > 256GB Titan Schwarz

Apple iPhone 16 Pro Max 256GB Titan Schwarz

4 Produktmeinungen: 5 ★★★★

Produktübersicht: 6,9 Zoll · Full HD · 48 MP · 8 GB RAM · 256 GB interner Speicher · iOS 18 · 4.685 mAh

Ähnliche Produkte: Phablet · 617 Apple Handys & Smartphones · 605 Apple LTE Smartphones · 357 Apple 5G Handys · K1 Handy · 5G Handy

Preisentwicklung

3M 6M 1J

Alle Varianten ab 1.109,90 €

256GB Titan Schwarz ab 1.114,90 €

512GB Titan Wüstensand Keine passenden Angebote

512GB Titan Schwarz Keine passenden Angebote

1TB Titan Weiß Keine passenden Angebote

25 N Ke Ar

Preiswecker stellen

Farbe: gold grau schwarz weiß

interner Speicher: 256 GB 512 GB 1.000 GB

Angebote: Neu ab 1.114,90 € B-Ware & Gebraucht ab 964,99 € In deiner Nähe

Top 10 Produkte 5G Handys

1 Apple iPhone 16 Pro ab 922,60 €

2 Samsung Galaxy S25 Ultra ab 848,90 €

3 Apple iPhone 16 ab 689,90 €

4 Samsung Galaxy S25 ab 578,90 €

5 Samsung Galaxy A56 ab 289,90 €

6 Apple iPhone 16 ab 559,90 €

Preisvergleich

Preis & Versand	Zahlungsarten*	Lieferung	Shop & Shopbewertung
1.114,90 € Günstiger Gesamtpreis 1.114,90 € inkl. Versand	PayPal VISA	Lieferung: bis Sa. 28.06. Rücksendung 14 Tage kostenfrei	ebay Marktplatz ★ 5,0 (47575) Verkauf durch: phones.com
1.139,00 € 1.145,90 € inkl. Versand Preis inkl. Gutschein	PayPal VISA Rechnung amazonepay	Lieferung: bis Mo. 16.06. Rücksendung 30 Tage kostenfrei	AS GOOD AS NEW ★ 4,8 (2587) Shop-Details
1.155,00 € 1.155,00 € inkl. Versand Preis inkl. Gutschein	PayPal VISA Rechnung Lastschrift Vorkasse	Lieferung: bis Sa. 21.06. Rücksendung 14 Tage kostenfrei	SO SMARTPHONE ONLY ★ 4,5 (446) Shop-Details

vast majority of offers are not mapped to product catalogue!

Relevancy Signals at idealo

idealo

- The users buy in shops
 - Purchase data is not directly observable at idealo.
- No "add to cart" like signals
 - Price alerts / wishlists instead

Samsung Galaxy A56 128GB Awesome Graphite

1 Produktmeinung: 0 ★★★★☆

Produktübersicht: 6,7 Zoll · Full HD · 50 MP · 8 GB RAM · 128 GB interner Speicher · Android 15 · 5.000 mAh [Produktdetails](#)

Ahnliche Produkte: Phablet · 1.439 Samsung Handys & Smartphones · [Android Handy · 5G Handy](#) · 693 Samsung 5G Handys · 1.367 Samsung Android Handys

Varianten: 128GB Awesome Graphite

Alle Varianten ab 308,49 €

128GB Awesome Graphite ab 309,90 €

Keine passenden Angebote

Enterprise Edition 128GB ...

256GB Awesome Graphite ab 319,90 €

Keine passenden Angebote

T2 OI Ke Ar

Preisentwicklung

3M 6M 1J

Preiswecker stellen

Farbe: grau, grün, pink, **schwarz**

interner Speicher: 128 GB, 256 GB

Angebote: Neu ab 309,90 €, B-Ware & Gebraucht ab 302,49 €, In deiner Nähe

Zuletzt angesehen: Seiko SNKL23J1 ab 176,32 €, Apple iPhone 16 Pro Max 256GB Titan Schwarz ab 1.114,90 €, Apple iPhone 16 Pro Max ab 1.109,90 €, ORIENT Bambino Automatic (RA-AC0) ab 158,40 €

Preisvergleich: Sofort lieferbar, ohne Rücksendekosten, Sortieren nach: Preis, Gesamtpreis

Angbotsbezeichnung: Preis & Versand, Zahlungsarten*, Lieferung, Shop & Shopbewertung

Samsung Galaxy A56 5G 128GB Awesome Graphite 309,90 € **Günstiger Gesamtpreis** 309,90 € inkl. Versand

PayPal VISA American Express Kreditkarte

Lieferung: bis Sa. 21.06. DHL smartport ★ 4,5 (497) Zum Shop

Samsung Galaxy A56 5G 128GB Awesome Graphite 309,90 € **Günstiger Gesamtpreis** 309,90 € inkl. Versand

PayPal VISA American Express Kreditkarte

Lieferung: bis Sa. 28.06. ebay ★ 4,9 (24313) Zum Shop

Rücksendung 30 Tage kostenfrei

Verkauf durch: smartport

smartport

Wonach suchst du?

Apple, Samsung, iPhone 16, iPhone 16e, iPhone 16 Pro, iPhone 16 Pro Max, Galaxy S24, Galaxy S25, Galaxy S25 Ultra, Galaxy A56

Startseite > Samsung Galaxy A56 5G 128GB Awesome ...

Samsung Galaxy A56 5G 128GB Awesome Graphite

Samsung | SKU: 211

Bestseller

★★★★★ 4,52 Sehr gut

€309,90

Neuware mit Herstellergarantie

Bezahle 103,30€/Monat in 3 Raten zu 0% Sollzinssatz mit Klarna.

Kostenloser Versand

Kostenlose Rücksendung innerhalb von 14 Tagen

Auf Lager

In den Warenkorb

Jetzt kaufen

Zahlungen & Sicherheit

Differenzbesteuert nach § 25a UStG. Die Mehrwertsteuer wird nicht separat ausgewiesen.

Before LTR: Keyword Search @ idealo



- Customized Lucene
- Manually tuned boosts
 - e.g. products over single offers
- Global item clicks and query-item clicks affect the boosts

	Bild	Details
Produkt id:203247542 Score : 173010 Clickouts (total):4359 Clickouts (query): 5359 Discount:0%		Apple iPhone 15 128GB Schwarz <small>+</small> Kategorien : Elektroartikel > Telekommunikation > Handys & Smartphones Product-Types : LTE Smartphone, 5G Handy, Phablet
Produkt id:203235721 Score : 164960 Clickouts (total):11660 Clickouts (query): 398 Discount:0%		Apple iPhone 15 <small>+</small> Kategorien : Elektroartikel > Telekommunikation > Handys & Smartphones Product-Types : LTE Smartphone, 5G Handy, Phablet
Produkt id:203247309 Score : 163070 Clickouts (total):1307 Clickouts (query): 819 Discount:0%		Apple iPhone 15 256GB Schwarz <small>+</small> Kategorien : Elektroartikel > Telekommunikation > Handys & Smartphones Product-Types : LTE Smartphone, 5G Handy, Phablet

Learning to Rank – The Problem

1	Apple TV 3		Product 0 3202627
2	LG C1 OLED TV		Product 0 202628669
3	LG OLED evo C3		Product 0 202668608
4	Apple iPod Classic 6G 160GB		Product 0 832742
5	LG OLED Flex 42LX3Q6LA		Product 0 202248350
6	LG G3 OLED		Product 0 202477155
7	Apple iTunes Guthaben		Product 0 2199358
8	LG 42LX3Q9LA		Product 0 202271959

Users
probably
want apple
tv

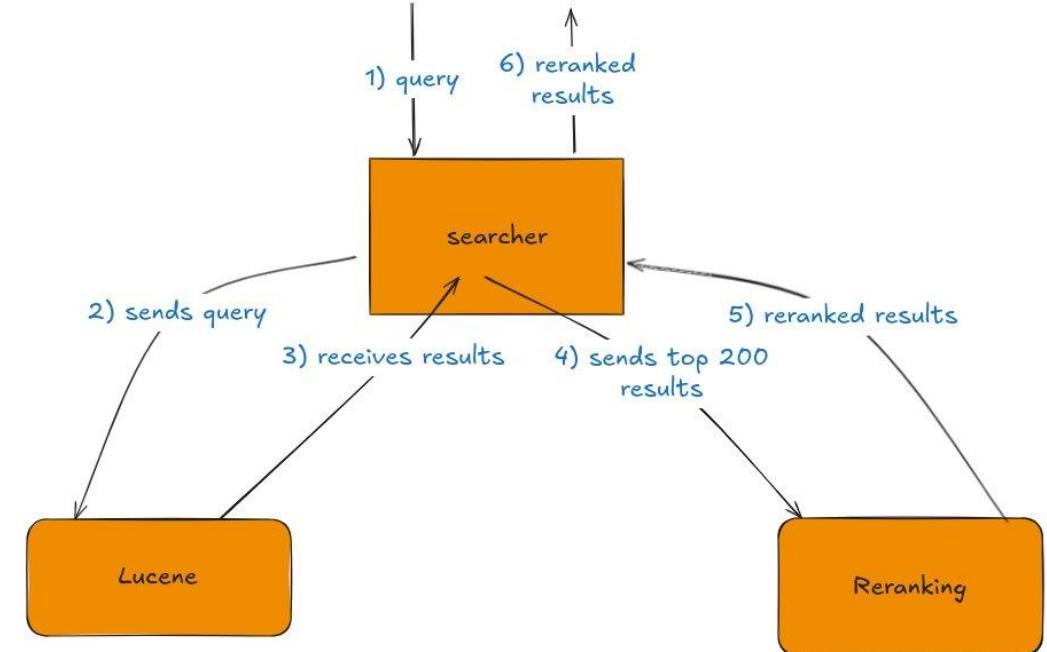
LTR Model



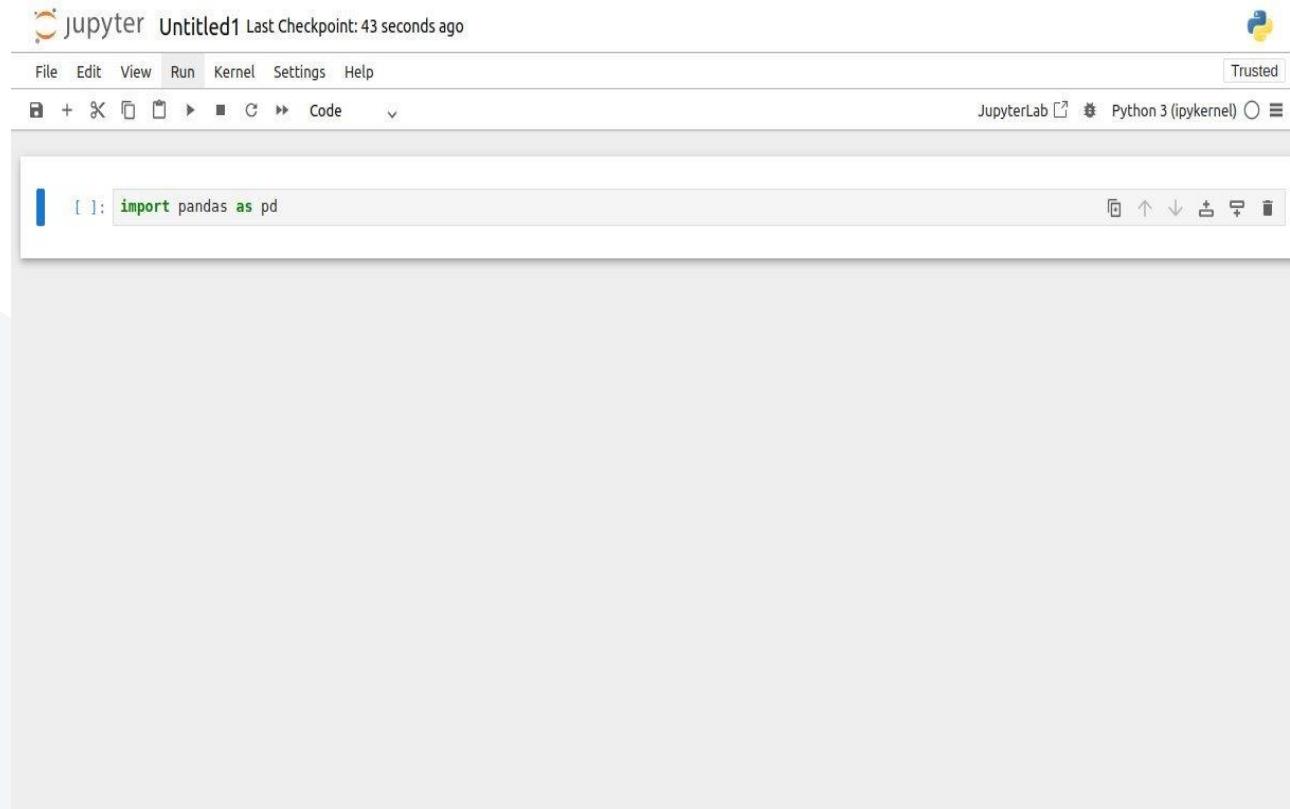
1	Apple TV 3		Product 0 3202627
2	Apple TV 4K 2022 (128 GB)		Product 0 202280976
3	Apple TV 4K 2022		Product 0 202280967
4	Apple TV 4K 2022 (64GB)		Product 0 202280953
5	Apple AirPods 3		Product 0 201554991
6	LG OLED evo C3		Product 0 202628669
7	Apple AirPods 3 mit Lightning-Case		Product 0 202729210
8	Apple TV 4K (2021)		Product 0 201231877

LTR Solution (ML reranking)

1. User sends query to a "searcher".
2. Searcher sends query to Lucene.
3. Receives results from Lucene.
4. Sends top 200 results to the LTR endpoint.
5. Receives reranked results from the endpoint.
6. Returns results to the user.



First steps. Setup a basic pipeline.

A screenshot of a Jupyter Notebook interface. The title bar says "jupyter Untitled1 Last Checkpoint: 43 seconds ago". The menu bar includes File, Edit, View, Run, Kernel, Settings, Help, and a "Trusted" badge. The toolbar below has icons for file operations like new, open, save, and run. The main area shows a code cell with the Python code "import pandas as pd" and a toolbar for cell execution. The status bar at the bottom says "JupyterLab" and "Python 3 (ipykernel)".

1. Collect user interactions (query-items clicks), including item features and query-item features.
2. Set labels based on user interactions (0-4 scale based on click-through rate - CTR).
3. Decide on the offline metric (we used NDCG).
4. Choose on simple features.
5. Conduct notebook experiments. Build the model (LambdaMART is a likely candidate).
6. Build the pipeline.
7. Run A/B tests.

Rules of machine learning:

<https://developers.google.com/machine-learning/guides/rules-of-ml>

Data

query	item id	Item position	text relevancy	clicks	impressions	ctr	category	price	relevance label
iphone	10	2	1000	369	1935	0.191	19116	100	4
iphone	20	1	1100	208	1915	0.109	19116	200	3
iphone	30	3	900	131	1887	0.069	19116	300	2
iphone	40	4	800	0	1935	0	2220	10	0
iphone x	70	2	5000	405	2165	0.187	19116	190	4
iphone x	80	1	5100	253	2152	0.118	19116	250	3
iphone x	90	3	4900	143	2119	0.067	19116	340	2
iphone x	100	4	3000	0	2119	0	2220	30	0



query-item features



item features

Challenge #1: Get better NDCG than the baseline.



Problem:

- **Clicks and labels are biased** toward the baseline ranking, introducing **position bias**.

Proposed solutions:

- You don't necessarily need a **much better NDCG** than the current baseline. It is more important what the model learned. But we still wanted to have it slightly better :).
- If you can get **LLM labeling**, you'll have an unbiased evaluation of labels.
- **Evaluating and incorporating bias** can be challenging.
- We didn't use **position features** for training, but maybe we should have.

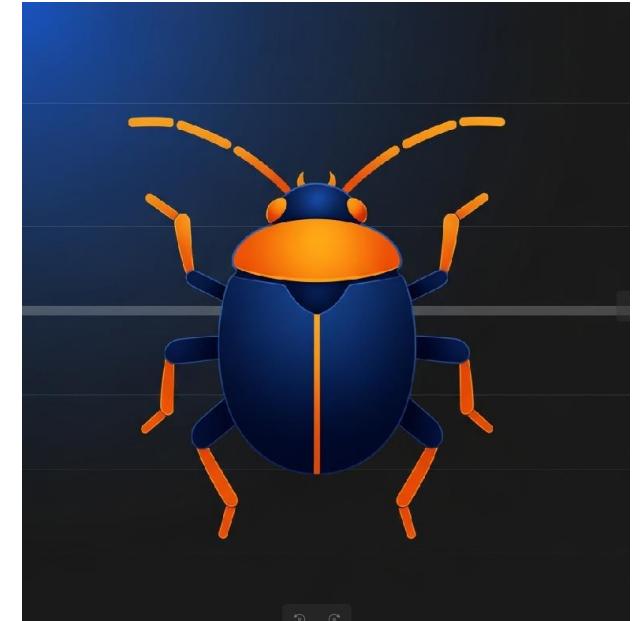


Challenge #2: Hidden bugs

Problem:

- There are many opportunities for bugs that do not produce errors but yield incorrect results.
- Incorrect metric implementation (for example, handling ties) and applied in a wrong way.
- Data leakage (taking features from the future).
- ...

Proposed solutions: tests, understand details :).



Use UI evaluation as well

idealo

Do UI side by side comparison to develop intuition about the model. Develop tools for easy debugging of the model.

Compare

Search Query: deuter 10l rucksack

Submit search

show item details show query click data

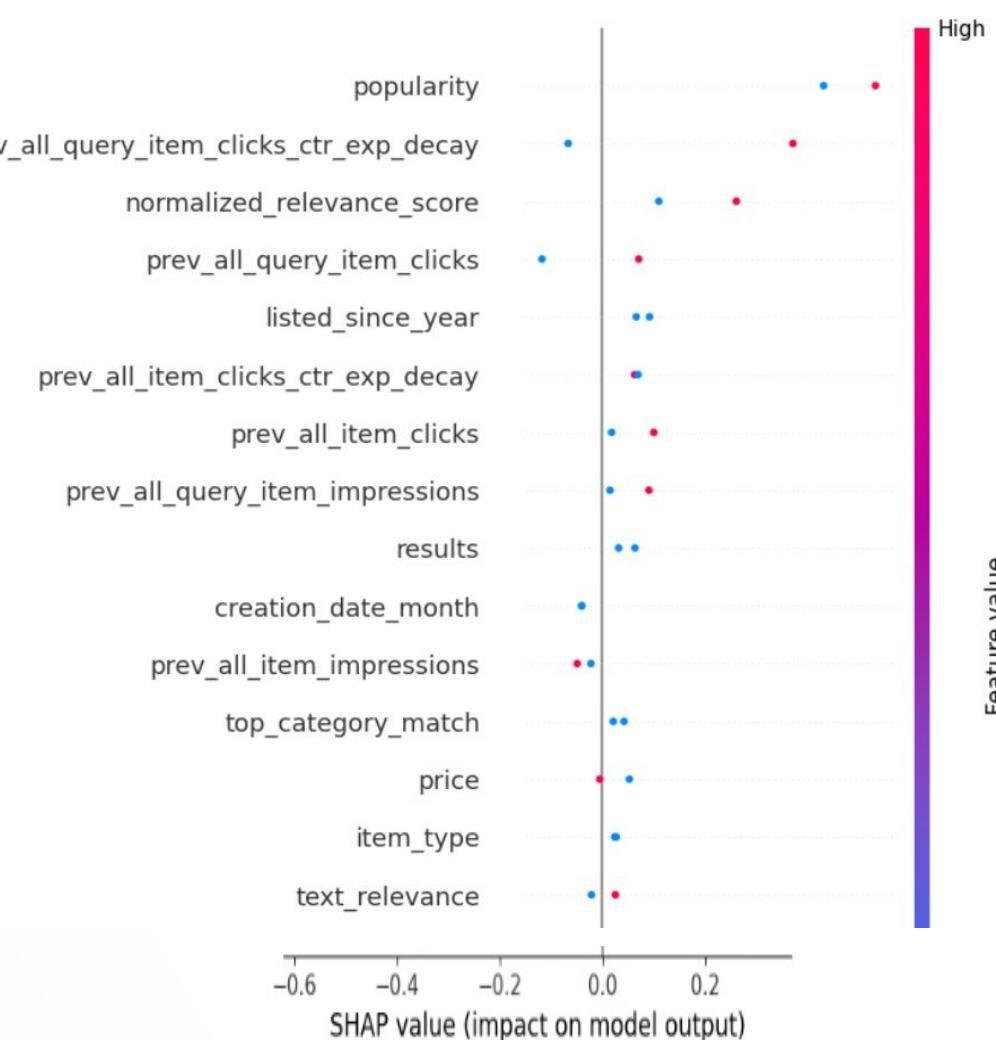
Search results

COMPARISON METRICS --> Jaccard Index: 0.622, RBO: 0.632

RESULT METRICS --> Tags: internal, ltr-reranked; Result Count: 205 (P: 166 / O: 39); Dispersion metrics: shop: 1.0, manut: 0.0;

position	type	item	image	position	type	item	image
1(0)	PRODUCT	Deuter Race Air (10l) (2021) id:201112988		1(0)	PRODUCT	Deuter Race Air (10l) (2021) id:201112988	
2(0)	PRODUCT	Deuter Race Air (10l) (2021) tin-shale id:201112989		2(+17)	PRODUCT	Deuter Waldtuchs 10 (2024) id:204790550	
3(0)	PRODUCT	Deuter Race Air (10l) (2021) azure-lapis id:201112991		3(+2)	PRODUCT	Deuter Waldtuchs 10 (2024) ruby/maroon id:204790573	
4(0)	PRODUCT	Deuter Race Air (10l) (2021) black id:20147114		4(-2)	PRODUCT	Deuter Race Air (10l) (2021) tin-shale id:201112989	
5(0)	PRODUCT	Deuter Waldtuchs 10 (2024) ruby/maroon id:204790573		5(-1)	PRODUCT	Deuter Race Air (10l) (2021) black id:20147114	

SHAP Image



red indicates high values and blue indicates low values

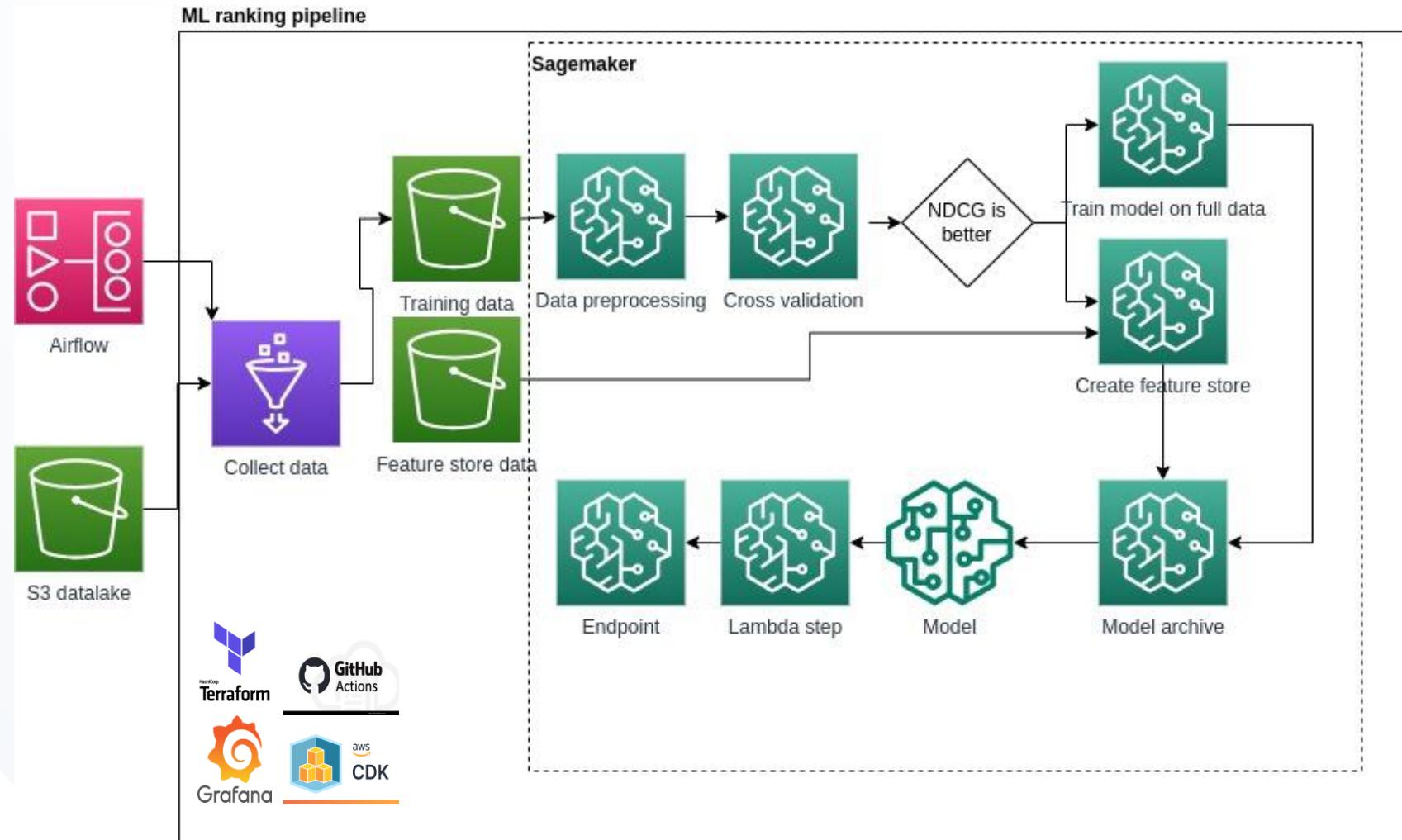
Challenge #3: Building the pipeline

Problem:

- Time-consuming setup.
While it provides immense convenience once established.
- Error-prone implementation.
- A lot of IaC code.

Proposed solution:

- If there's nothing to update during an A/B test, skip this step until you get a good A/B test result.



Challenge #4: The same model in production as during training

Problem:

- Discrepancies between experimental and inference code.
- Incorrect feature passing to inference endpoints.

Proposed solutions:

- Ensure training-serving parity.
- Make sure you can debug what features being passed to your model.
- NDCG evaluation across the entire search.

Search Toolbox

Home Debug Offline Evaluation Expert **Offline evaluation** Search Version Comparison Result-CSV visualization

select analysis type

select golden query

Endpoint

Test ID (pattern: sam-1234)

Variants A B C D E F G

Sample Size

Custom suffix

Preheat Scales Jarvis and Routers before test run (takes ~ a minute)

Preheat LTR Endpoints Scale rank-end (Variant A)
 Scale rank-end-test (Variant B)
 Scale rank-end-dev (Variant C)

Please select test variants you want to evaluate. Make sure to fill in fields above before editing requests manually.

```
{
  "clientId": "offline-evaluation",
  "count": 200,
  "includeProductDetails": "true",
  "locale": "de_DE",
  "offset": 0,
  "query": "",
  "testVariants": {},
  "sessionInfo": {},
  "blockedModifiers": [],
  "mode": "WEB",
}
```

```
{
  "clientId": "offline-evaluation",
  "count": 200,
  "includeProductDetails": "true",
  "locale": "de_DE",
  "offset": 0,
  "query": "",
  "testVariants": {},
  "sessionInfo": {},
  "blockedModifiers": [],
  "mode": "WEB",
}
```

Challenge #5: Differentiating click data in A/B test

Problem:

- Distinguish clicks from new ML model vs. static ranking during A/B test.
- Achieving this distinction was **difficult** for us.

Proposed solutions:

- When an A/B test for a new ranking model begins, **clicks generated by the *baseline* (control) ranking system should no longer be considered.**
- If not possible, **check for data leakage** effects.
- It is better to avoid retraining during A/B test.



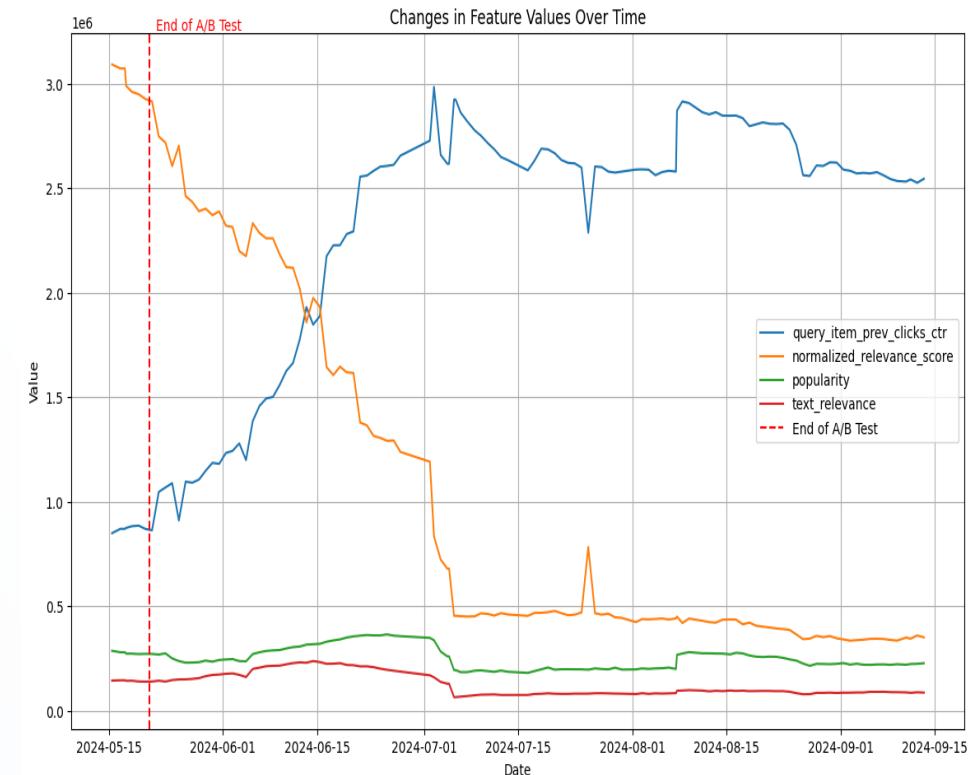
Challenge #6: Deployed in production. How to retrain?

The feedback loop problem:

- Model performance degrades when it's trained on implicit feedback (e.g., clicks) that was generated by its own previous rankings.
- Initial biases are amplified over time.

Potential solutions: breaking the cycle

- Generate unbiased labels with LLM.
- Implement exploration traffic.
- Apply bias correction.



Changes of features importance since the model deployed in production. Text relevance feature importance dropped significantly. Previous query items CTR feature importance increased drastically.

Challenge #7: What to monitor?

Besides technical metrics like latency and errors, we suggest to **track overall click-through rate (CTR)** at the beginning. Then many fancy things like concept drift, data drift and so on.

Important: Avoid tracking differences in NDCG between a baseline ranking and your new ranking. This can be misleading because user clicks are inherently biased towards the existing production model.



Challenge #8: Latency

- **Prioritize vectorized operations:**

- Vectorized (fast):

```
df["ratio"] = 100 * (df["x"] /  
df["y"])
```

- Row-wise (slow):

```
df["ratio"] =  
df.apply(calculate_ratio)
```

- **LightGBM & high cardinality categorical features:** treat high-cardinality categorical features (2K categories, for example) as numerical for improved performance.
- **Solution:** validate performance, conduct thorough benchmark and load tests.

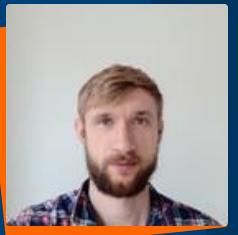


AB test metric	% Change
Consolidated CR	+1,37%
Ratio of sessions exiting after the 1st MainSearch	-2,32%
Exit rate	-1,63%

- Average RPS: **60**
- Average latency: **40ms**
- Model: custom image of **Lambdarank Lightgbm**
- Training data size: **5,5mio query items pairs**
- Number of features: **46**
- Pipeline running time (without finetuning): **1,5 hours**
- We rerank top 200 results.
- We apply reranking on the whole catalog.



Questions



Gennady Shabanov

Machine Learning Engineer



Atakan Filgöz

Machine Learning Engineer