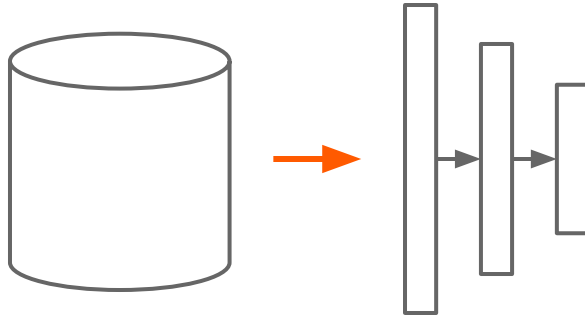
The background of the slide is a photograph of a group of students in a library or study hall, overlaid with a semi-transparent orange filter. The students are gathered around a table, looking at papers and books, suggesting a collaborative learning environment.

How to use Machine Learning models to build training sets?

Piotr Rybak

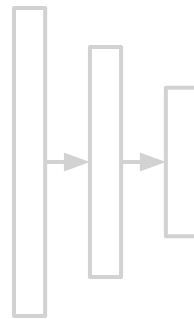
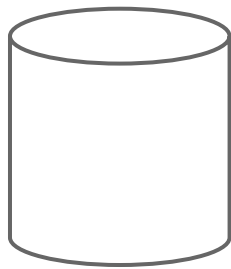
allegro

Dataset

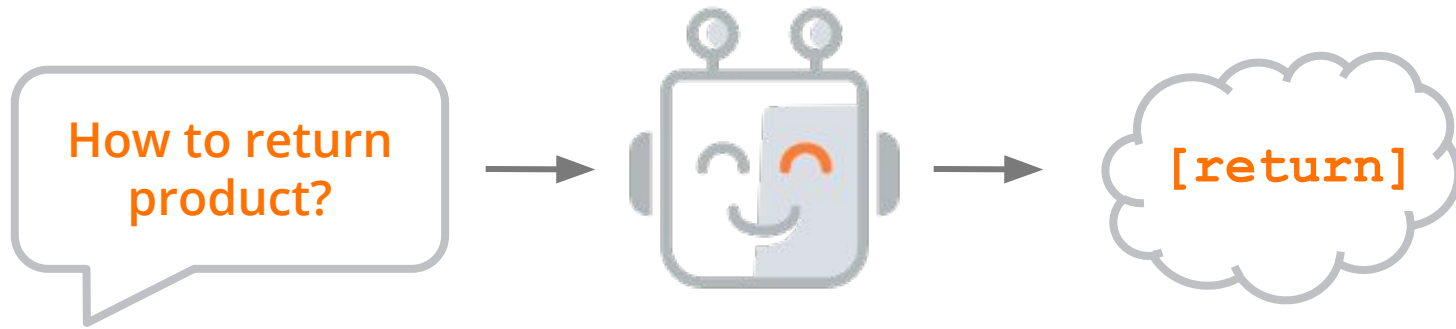


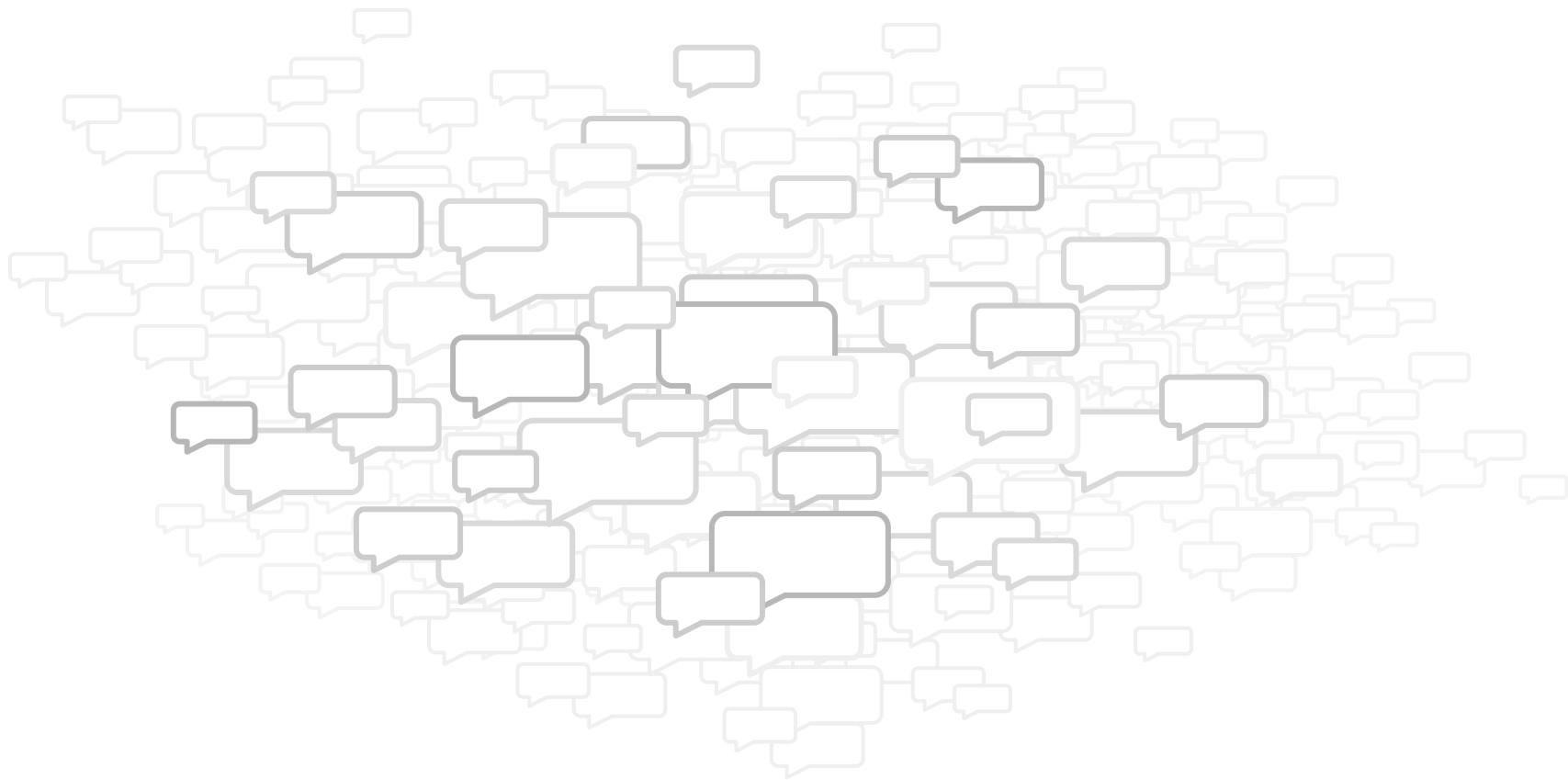
ML Model

Dataset



ML Model







[return status]

[how to return]

[change delivery]

[delivery status]

[account number]

...



[return status]

[how to return]

[change delivery]

[delivery status]

[account number]

...

> Difficult task

> Time-consuming

> Low recall

> Low consistency (~50% for 100 classes)



[return status]

[how to return]

[change delivery]

[delivery status]

[account number]

...

> Difficult task

> Time-consuming

> Low recall

> Low consistency (~50% for 100 classes)

> Class imbalance



[return status]
[how to return]
[change delivery]
[delivery status]
[account number]

...

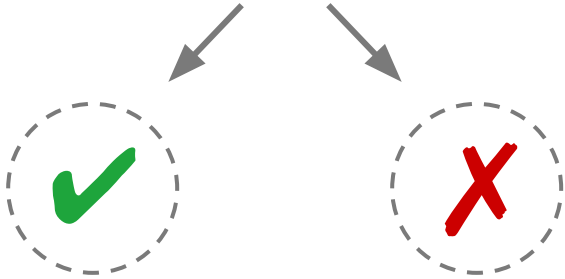
> **Difficult task**

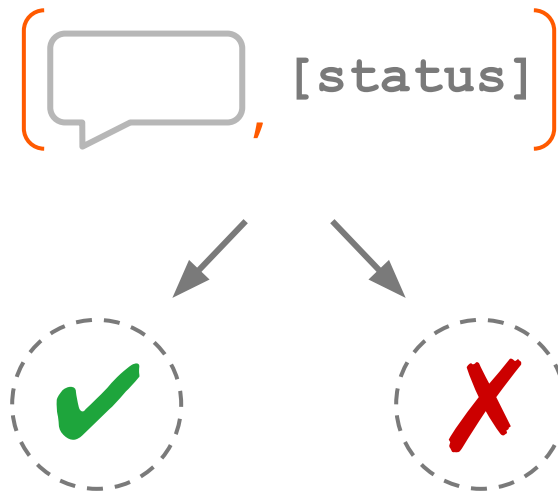
- > Time-consuming
- > Low recall
- > Low consistency (~50% for 100 classes)

> **Class imbalance**

> **Fixed set of classes**

 [status]

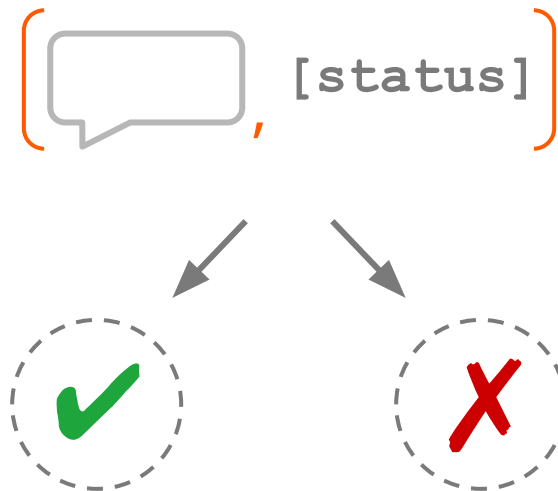




> **Easy task**

> Very fast

> High consistency (~95% for each class)

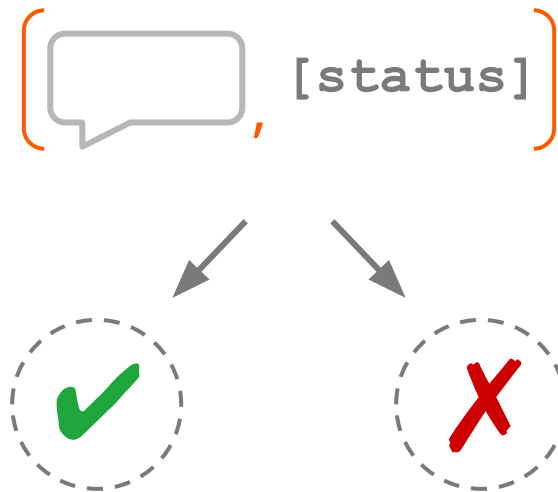


> **Easy task**

> Very fast

> High consistency (~95% for each class)

> **As much examples as needed**



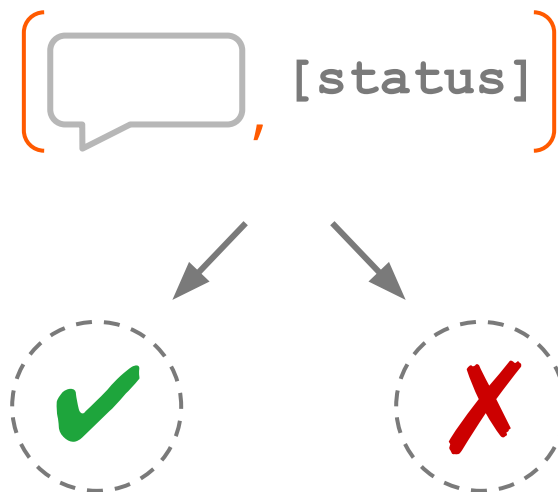
> **Easy task**

> Very fast

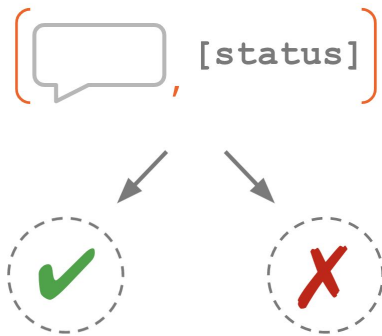
> High consistency (~95% for each class)

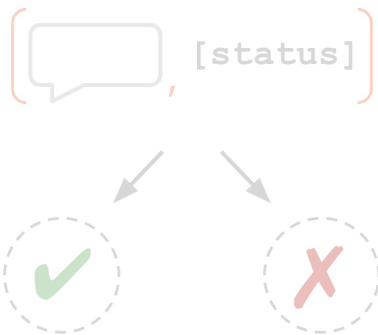
> **As much examples as needed**

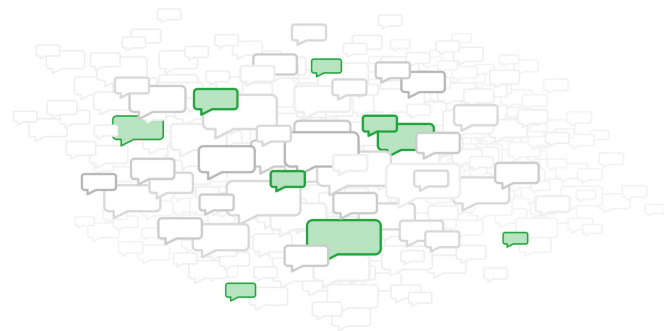
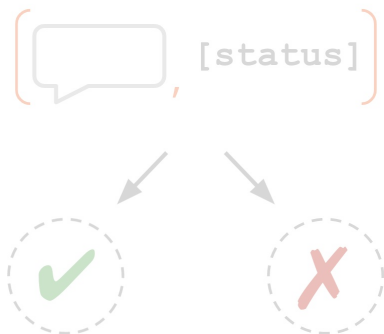
> **Can introduce new classes**

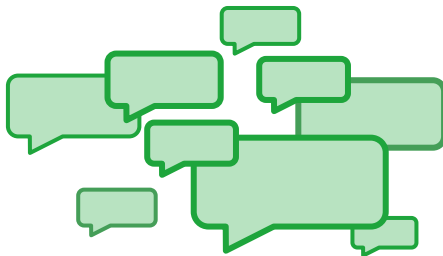
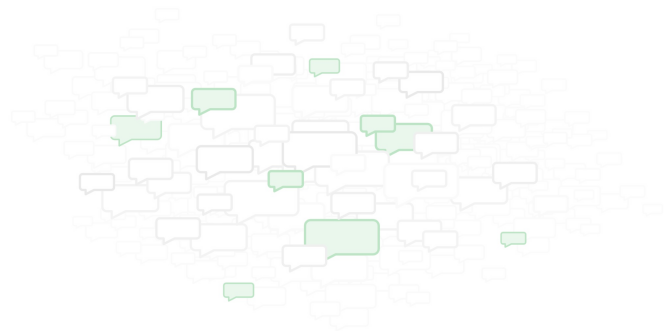
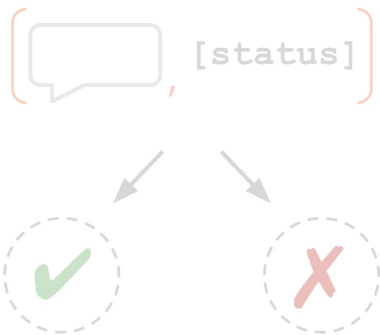


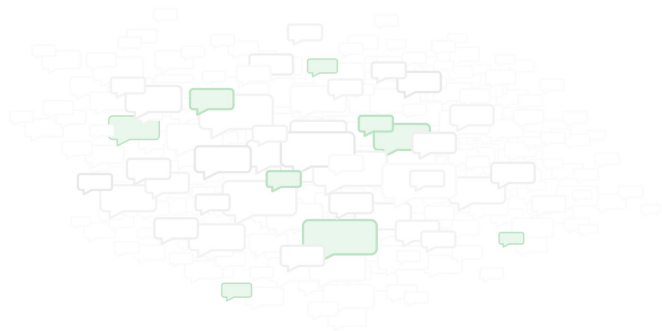
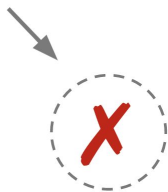
- > **Easy task**
 - > Very fast
 - > High consistency (~95% for each class)
- > **As much examples as needed**
- > **Can introduce new classes**
- > **Wastes a lot of time**
 - > You mostly annotate negatives

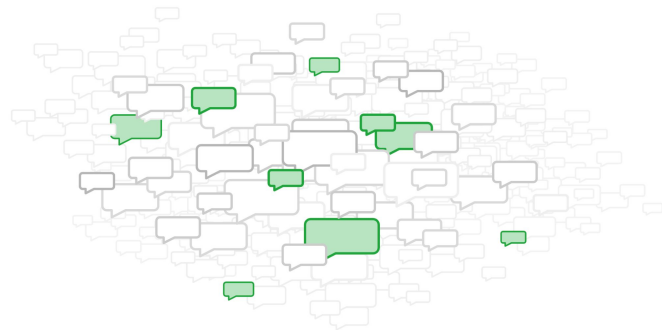
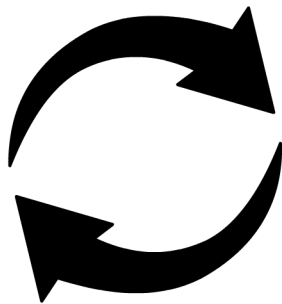
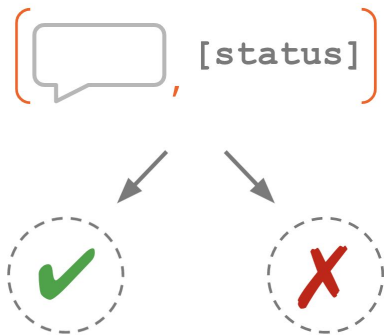


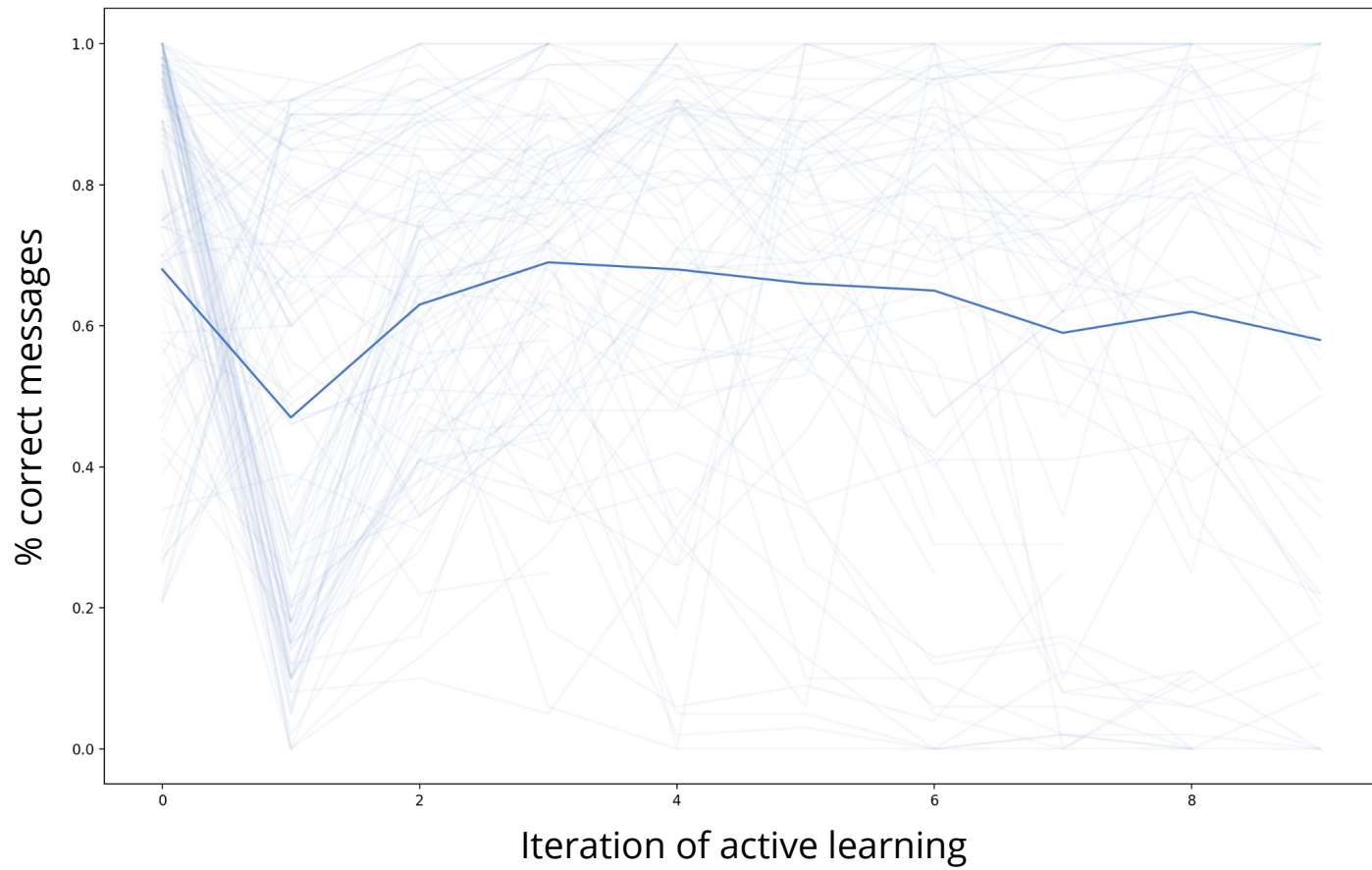


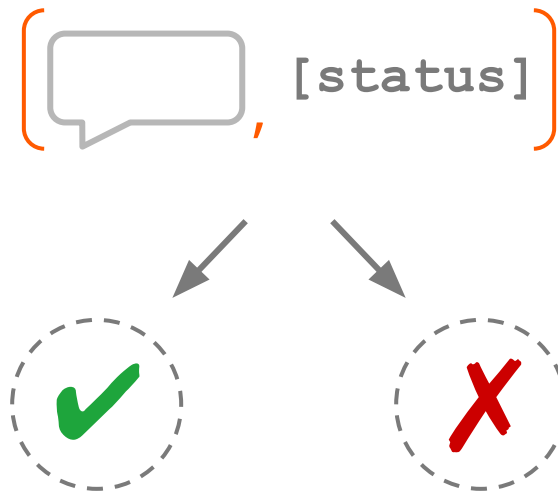












- > **Easy task**
 - > Very fast
 - > High consistency (~95% for each class)
- > **As much examples as needed**
- > **Can introduce new classes**
- > ~~Wastes a lot of time~~ **Biased sample**

Spurious correlations

How to return a dress?

→ **[how to return]**

How to return to the main page?

→ **[how to return]**

Spurious correlations

How to return a dress?

→ [how to return]

How to return to the main page?

→ [how to return]

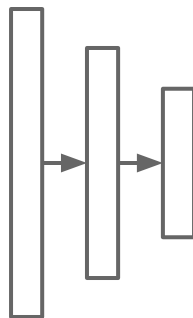
How to return a dress?

→ [how to return]

How to buy a dress?

→ [how to return]

Intent features



... exchange ...

... return ...

How do I ...

How to return ...

... dress

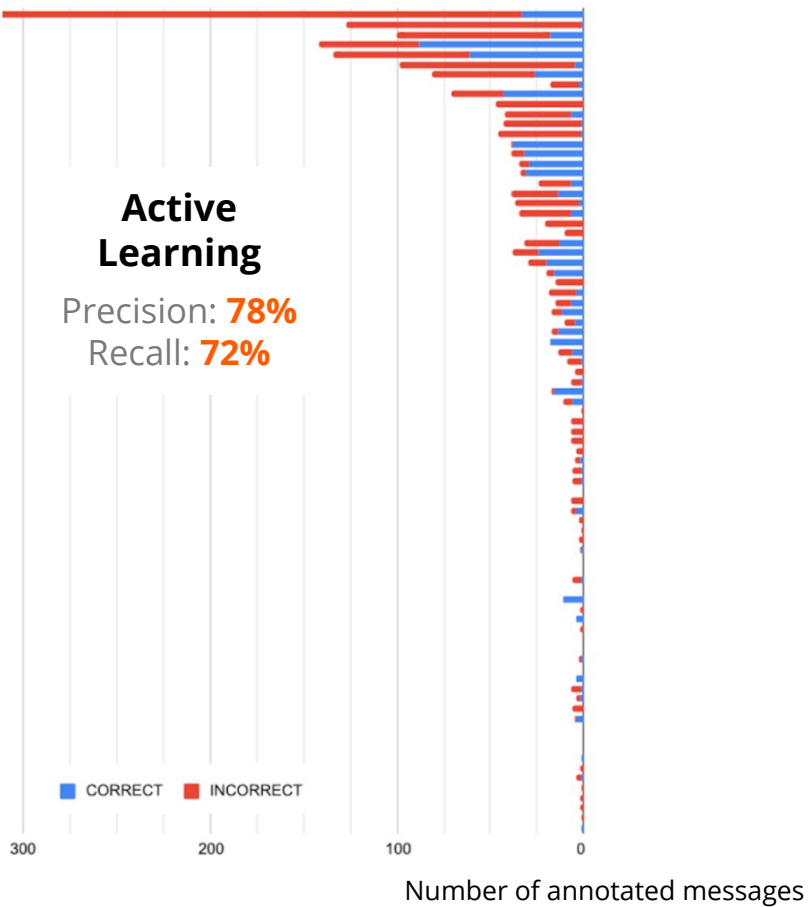
... returning ...

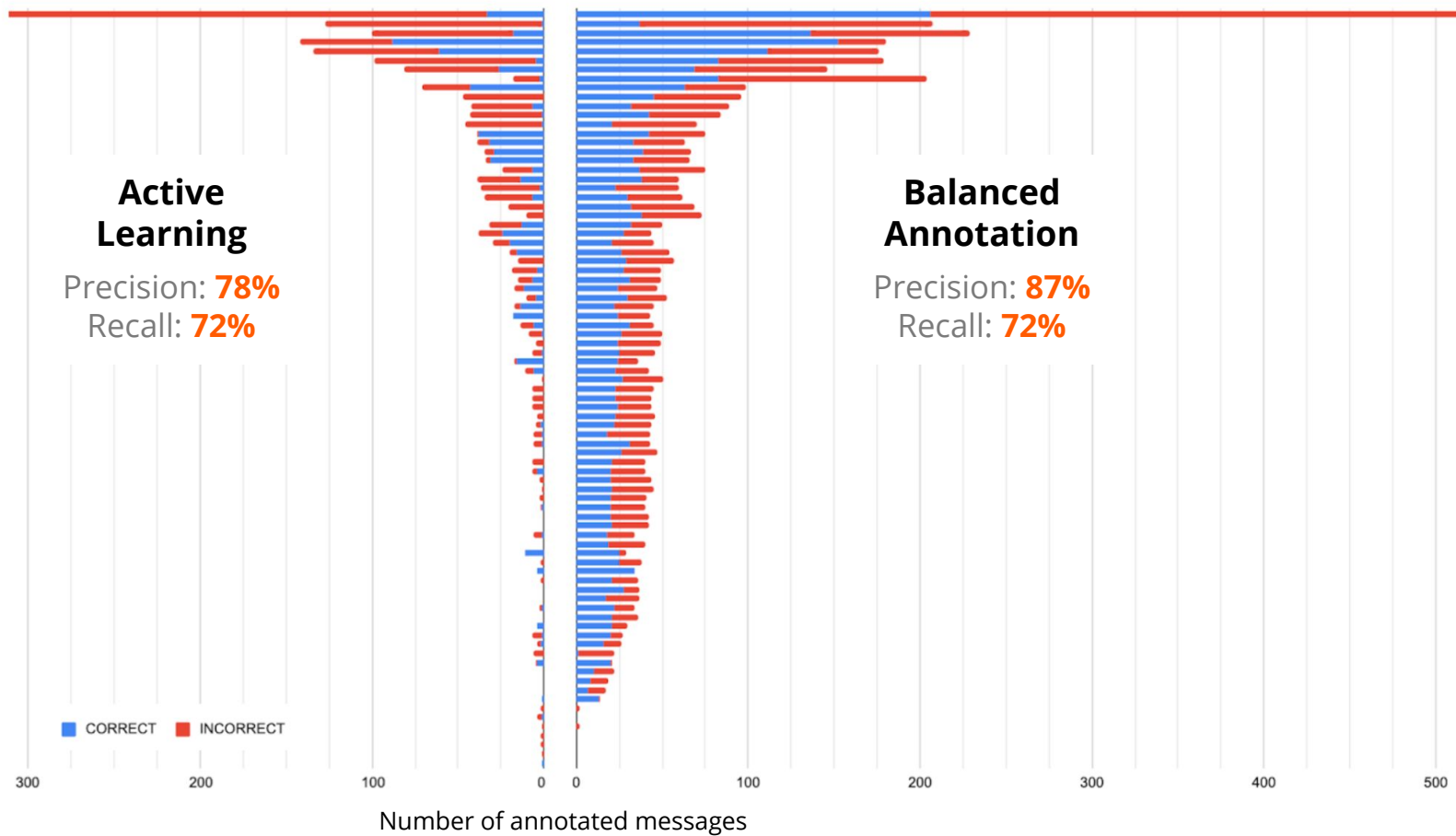
Where can I return ...

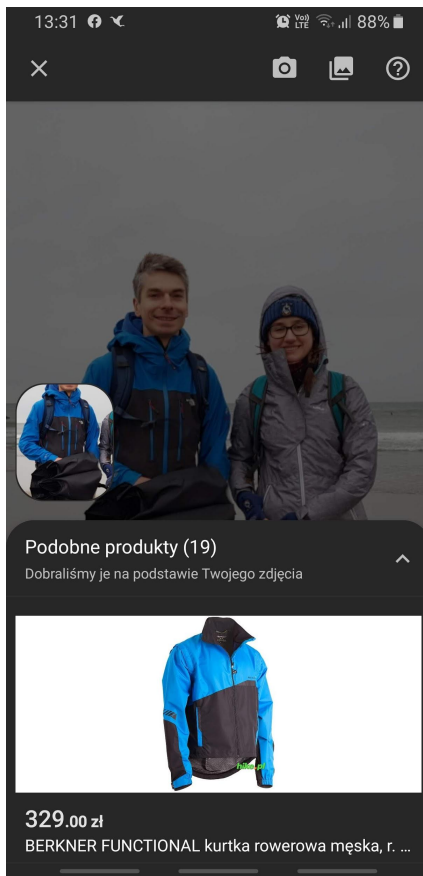
Active Learning

Precision: 78%

Recall: 72%







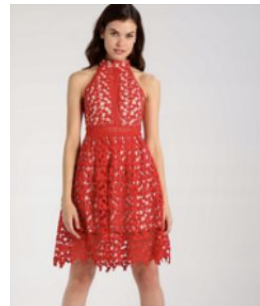
Metric Learning: Triplet-loss



Metric Learning: Triplet-loss



Metric Learning: Triplet-loss



Metric Learning: Triplet-loss



Active learning (?)

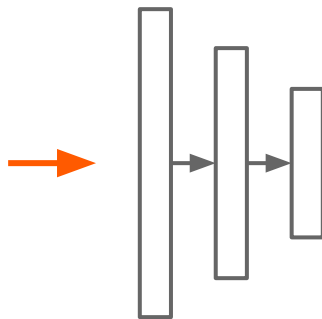


User query

Active learning (?)



User query

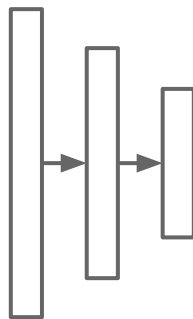


ResNet

Active learning (?)



User query

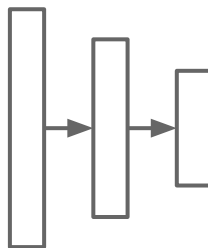


ResNet



"Similar" images

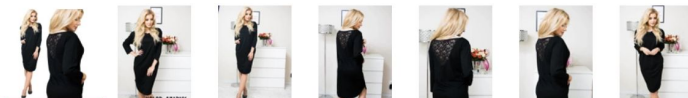
ImageNet
ResNet



Accuracy@5

2%

Random pairs from the same item



☆ OBSERWUJ

ZMYŚLOWA duża SUKIENKA z koronką 3XL 4XL 46 48

od [womanonline](#)

98,7% poleca sprzedawcę

49,90 zł SMART!

30 osób kupiło 30 sztuk

Darmowa dostawa

Najszybsza dostawa: wt. 24 wrz. - śr. 25 wrz. w punkcie

Darmowy zwrot

SZCZEGÓŁY DOSTAWY

Liczba sztuk

— 1 + z 9970 sztuk

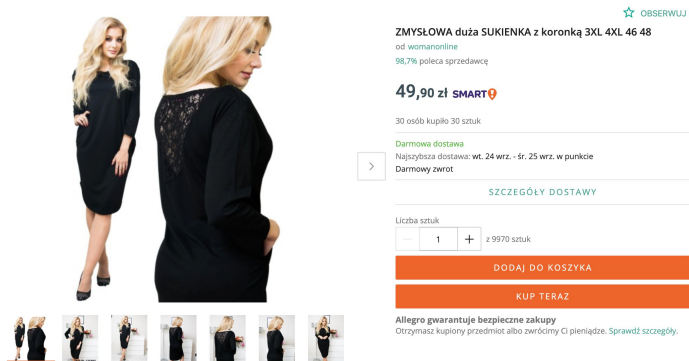
DODAJ DO KOSZYKA

KUP TERAZ

Allegro gwarantuje bezpieczne zakupy

Otrzymasz kupiony przedmiot albo zwrócimy Ci pieniądze. [Sprawdź szczegóły.](#)

Random pairs from the same item



Incorrect pairs



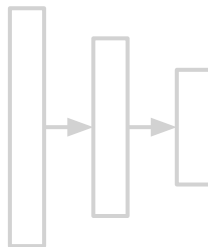
	S	M	L	XL
BIUST:	84 cm (max)	86 cm (max)	92 cm (max)	98 cm (max)
TALIA:	68 cm (max)	70 cm (max)	76 cm (max)	80 cm (max)
BIODRA:	86 cm (max)	94 cm (max)	100 cm (max)	103 cm (max)



Trivial pairs



ImageNet
ResNet



Accuracy@5

2%

Random
Pairs



6%

Domain mismatch



User query



Product image

Desired image types

Product-shop



Model-shop



Product-street



Model-street



Redundant image types

Multiple



Detail



Trash

Accept Customize clothes , you just need send me your pattern
Both women and men can wear

Upper hoodless sweatshirt-sleeved size table Unit: CM (cm)

Size	Chest	Length
XS	42	65
X	46	68
S	48	67
M	52	70
L	56	73
XL	60	76
XXL	64	79
XXXL	68	82

Upper hoodless sweatshirt-sleeved size recommended table Unit: CM (cm)

Size	XS	X	S	M	L	XL	XXL	XXXL	XXXL
150	XS	XS	S	S	M	L	XL	XXL	XXXL
160	XS	XS	S	S	M	L	XL	XXL	XXXL
170	XS	XS	S	S	M	L	XL	XXL	XXXL
180	XS	XS	S	S	M	L	XL	XXL	XXXL
190	XS	XS	S	S	M	L	XL	XXL	XXXL
200	XS	XS	S	S	M	L	XL	XXL	XXXL
210	XS	XS	S	S	M	L	XL	XXL	XXXL
220	XS	XS	S	S	M	L	XL	XXL	XXXL
230	XS	XS	S	S	M	L	XL	XXL	XXXL
240	XS	XS	S	S	M	L	XL	XXL	XXXL
250	XS	XS	S	S	M	L	XL	XXL	XXXL
260	XS	XS	S	S	M	L	XL	XXL	XXXL
270	XS	XS	S	S	M	L	XL	XXL	XXXL
280	XS	XS	S	S	M	L	XL	XXL	XXXL
290	XS	XS	S	S	M	L	XL	XXL	XXXL
300	XS	XS	S	S	M	L	XL	XXL	XXXL
310	XS	XS	S	S	M	L	XL	XXL	XXXL
320	XS	XS	S	S	M	L	XL	XXL	XXXL
330	XS	XS	S	S	M	L	XL	XXL	XXXL
340	XS	XS	S	S	M	L	XL	XXL	XXXL
350	XS	XS	S	S	M	L	XL	XXL	XXXL
360	XS	XS	S	S	M	L	XL	XXL	XXXL
370	XS	XS	S	S	M	L	XL	XXL	XXXL
380	XS	XS	S	S	M	L	XL	XXL	XXXL
390	XS	XS	S	S	M	L	XL	XXL	XXXL
400	XS	XS	S	S	M	L	XL	XXL	XXXL
410	XS	XS	S	S	M	L	XL	XXL	XXXL
420	XS	XS	S	S	M	L	XL	XXL	XXXL
430	XS	XS	S	S	M	L	XL	XXL	XXXL
440	XS	XS	S	S	M	L	XL	XXL	XXXL
450	XS	XS	S	S	M	L	XL	XXL	XXXL
460	XS	XS	S	S	M	L	XL	XXL	XXXL
470	XS	XS	S	S	M	L	XL	XXL	XXXL
480	XS	XS	S	S	M	L	XL	XXL	XXXL
490	XS	XS	S	S	M	L	XL	XXL	XXXL
500	XS	XS	S	S	M	L	XL	XXL	XXXL
510	XS	XS	S	S	M	L	XL	XXL	XXXL
520	XS	XS	S	S	M	L	XL	XXL	XXXL
530	XS	XS	S	S	M	L	XL	XXL	XXXL
540	XS	XS	S	S	M	L	XL	XXL	XXXL
550	XS	XS	S	S	M	L	XL	XXL	XXXL
560	XS	XS	S	S	M	L	XL	XXL	XXXL
570	XS	XS	S	S	M	L	XL	XXL	XXXL
580	XS	XS	S	S	M	L	XL	XXL	XXXL
590	XS	XS	S	S	M	L	XL	XXL	XXXL
600	XS	XS	S	S	M	L	XL	XXL	XXXL
610	XS	XS	S	S	M	L	XL	XXL	XXXL
620	XS	XS	S	S	M	L	XL	XXL	XXXL
630	XS	XS	S	S	M	L	XL	XXL	XXXL
640	XS	XS	S	S	M	L	XL	XXL	XXXL
650	XS	XS	S	S	M	L	XL	XXL	XXXL
660	XS	XS	S	S	M	L	XL	XXL	XXXL
670	XS	XS	S	S	M	L	XL	XXL	XXXL
680	XS	XS	S	S	M	L	XL	XXL	XXXL
690	XS	XS	S	S	M	L	XL	XXL	XXXL
700	XS	XS	S	S	M	L	XL	XXL	XXXL
710	XS	XS	S	S	M	L	XL	XXL	XXXL
720	XS	XS	S	S	M	L	XL	XXL	XXXL
730	XS	XS	S	S	M	L	XL	XXL	XXXL
740	XS	XS	S	S	M	L	XL	XXL	XXXL
750	XS	XS	S	S	M	L	XL	XXL	XXXL
760	XS	XS	S	S	M	L	XL	XXL	XXXL
770	XS	XS	S	S	M	L	XL	XXL	XXXL
780	XS	XS	S	S	M	L	XL	XXL	XXXL
790	XS	XS	S	S	M	L	XL	XXL	XXXL
800	XS	XS	S	S	M	L	XL	XXL	XXXL
810	XS	XS	S	S	M	L	XL	XXL	XXXL
820	XS	XS	S	S	M	L	XL	XXL	XXXL
830	XS	XS	S	S	M	L	XL	XXL	XXXL
840	XS	XS	S	S	M	L	XL	XXL	XXXL
850	XS	XS	S	S	M	L	XL	XXL	XXXL
860	XS	XS	S	S	M	L	XL	XXL	XXXL
870	XS	XS	S	S	M	L	XL	XXL	XXXL
880	XS	XS	S	S	M	L	XL	XXL	XXXL
890	XS	XS	S	S	M	L	XL	XXL	XXXL
900	XS	XS	S	S	M	L	XL	XXL	XXXL
910	XS	XS	S	S	M	L	XL	XXL	XXXL
920	XS	XS	S	S	M	L	XL	XXL	XXXL
930	XS	XS	S	S	M	L	XL	XXL	XXXL
940	XS	XS	S	S	M	L	XL	XXL	XXXL
950	XS	XS	S	S	M	L	XL	XXL	XXXL
960	XS	XS	S	S	M	L	XL	XXL	XXXL
970	XS	XS	S	S	M	L	XL	XXL	XXXL
980	XS	XS	S	S	M	L	XL	XXL	XXXL
990	XS	XS	S	S	M	L	XL	XXL	XXXL
1000	XS	XS	S	S	M	L	XL	XXL	XXXL

Hand measurement will have discrepancy of about 1-2cm. T-shirt Size
It is Asia size, please don't choose it as a US/EU/UK size. It is smaller than US/AU/EU size.

Other



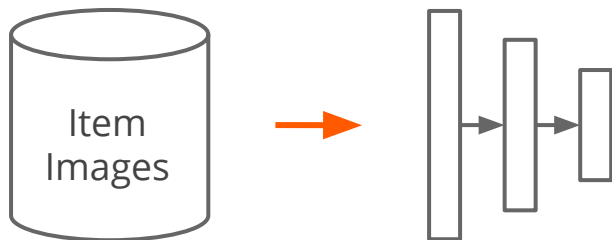
Manually annotate image types



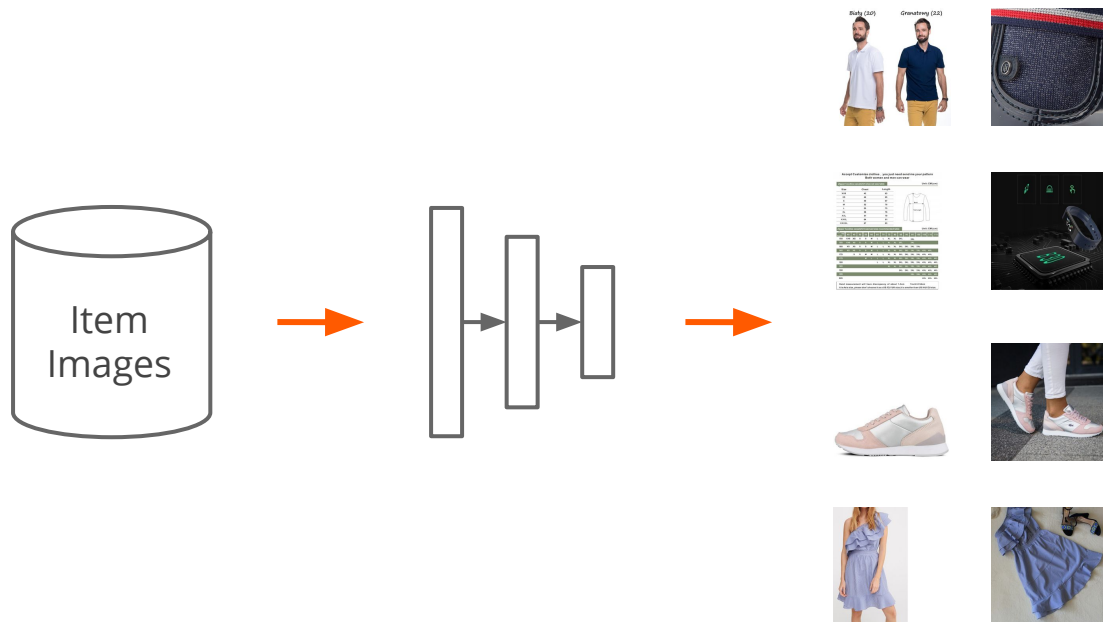
Automatic dataset creation



Automatic dataset creation



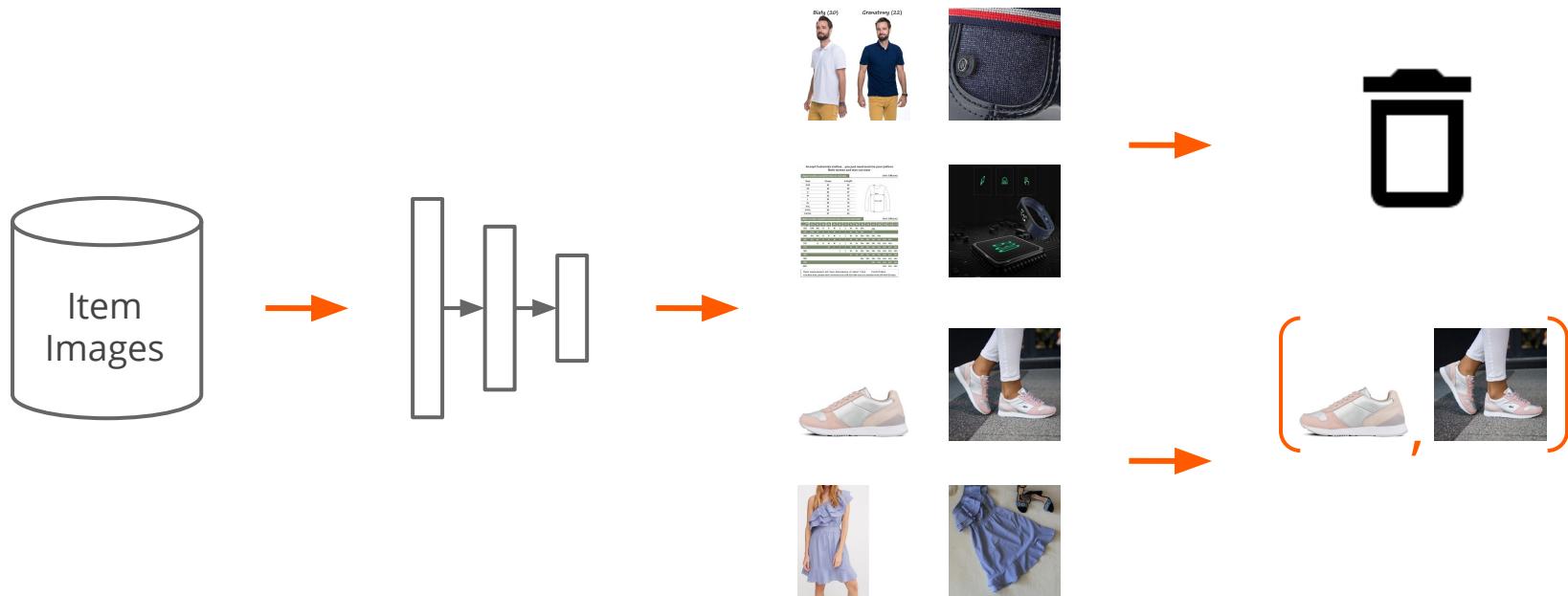
Automatic dataset creation



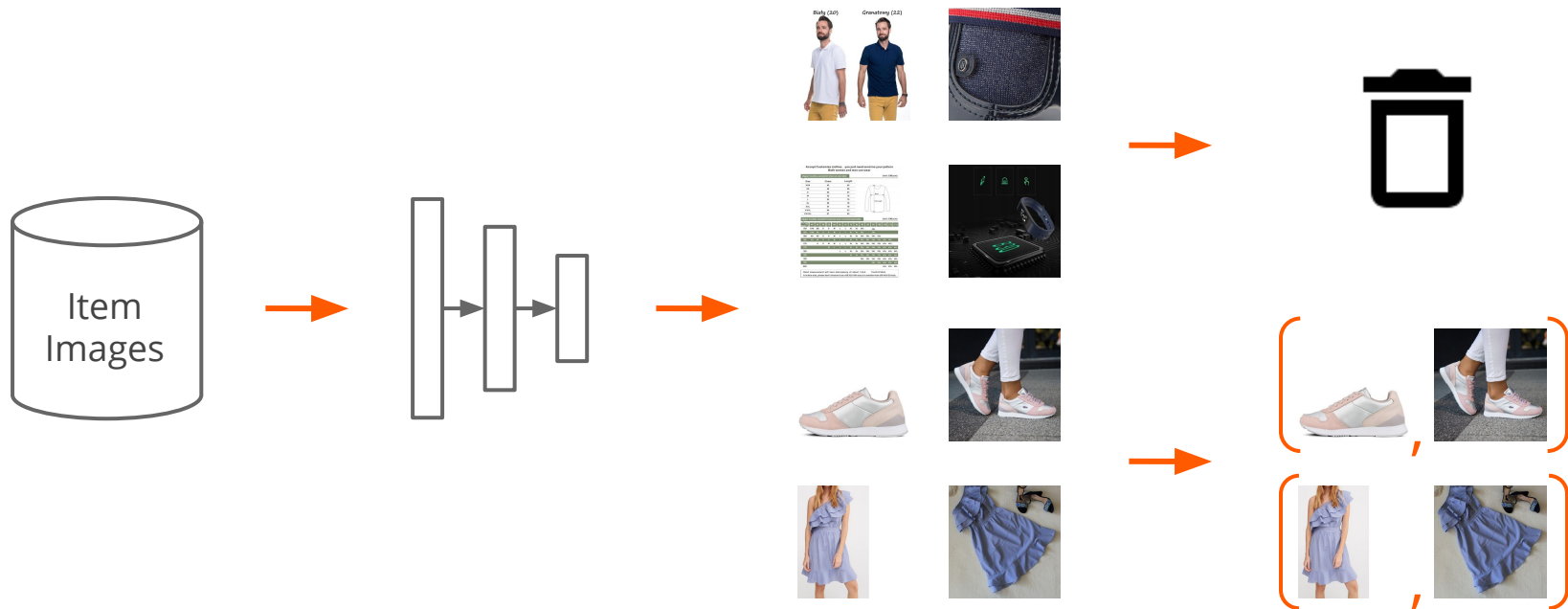
Automatic dataset creation



Automatic dataset creation

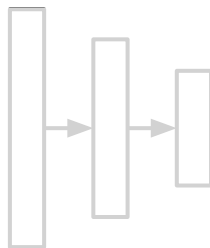


Automatic dataset creation



Accuracy@5

ImageNet
ResNet



2%

Random
Pairs



6%

Filtered
Pairs



50%

Main takeaways

- > Automate your work using ML

Main takeaways

- > Automate your work using ML
- > Avoid spurious correlations by balancing features

Main takeaways

- > Automate your work using ML
- > Avoid spurious correlations by balancing features
- > Be creative!



THANKS!

piotr.rybak@allegro.pl

We are looking for
Research Engineer

- Interesting and challenging projects (NLP, RL, Vision, Ranking)
- 20% of time for pure research
- Weekly seminars

Learn more: smrtr.io/63WHX

Our site: ml.allegro.tech

allegro