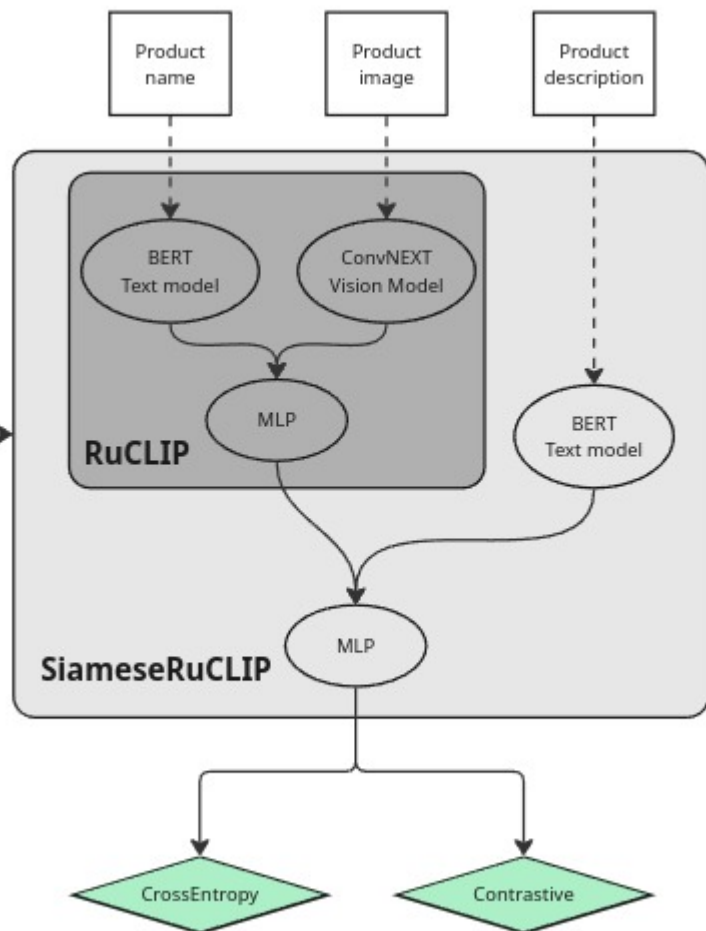
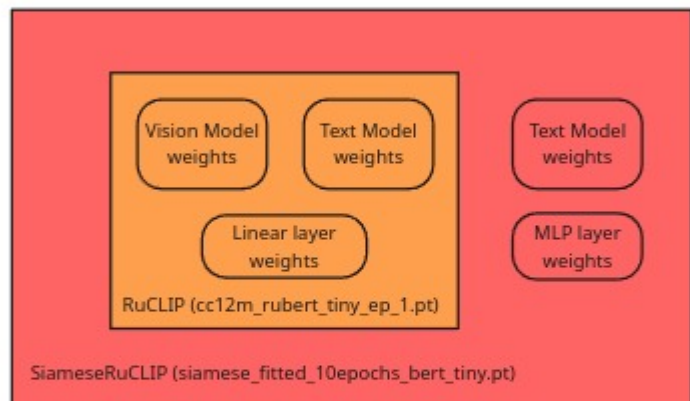
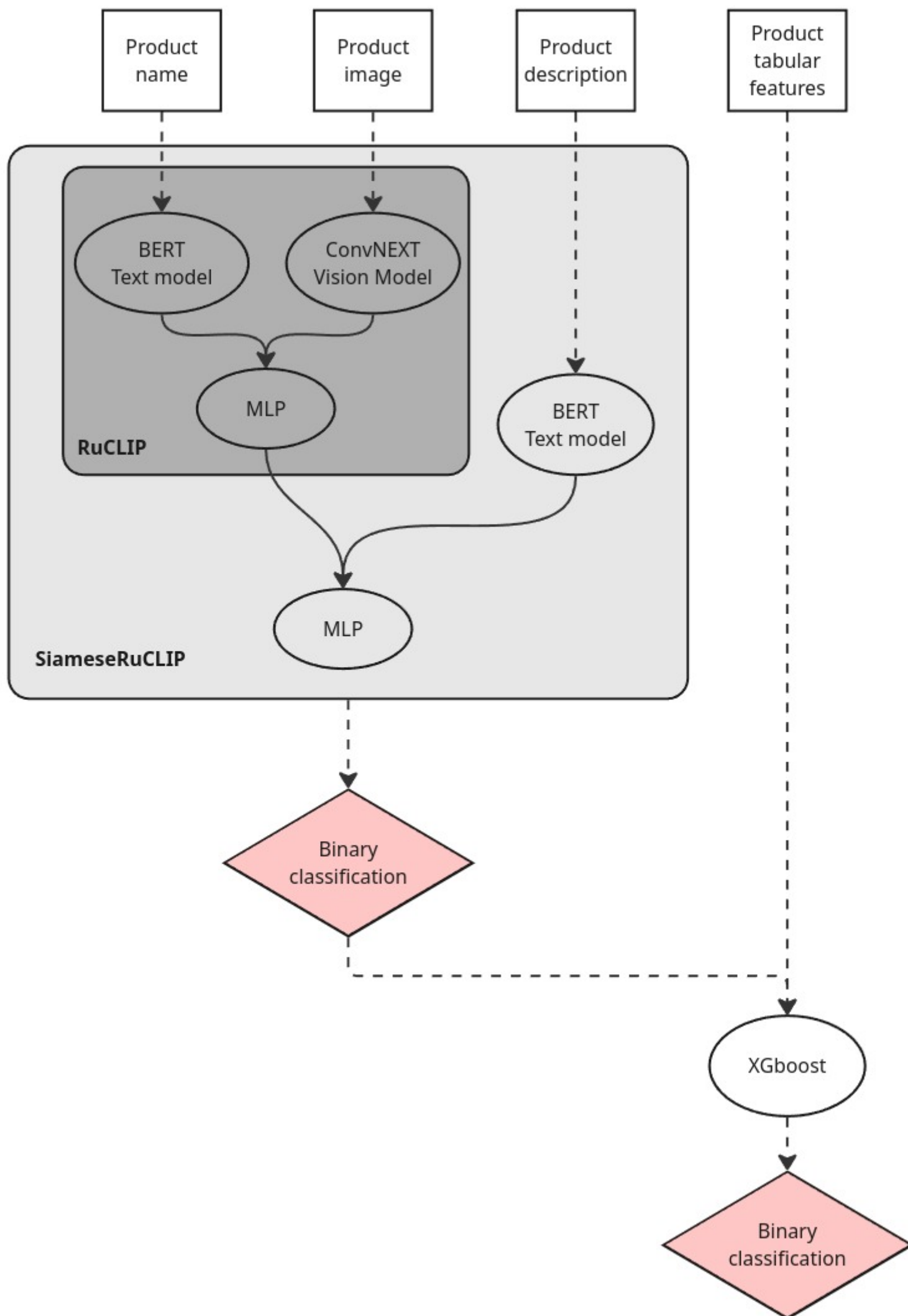
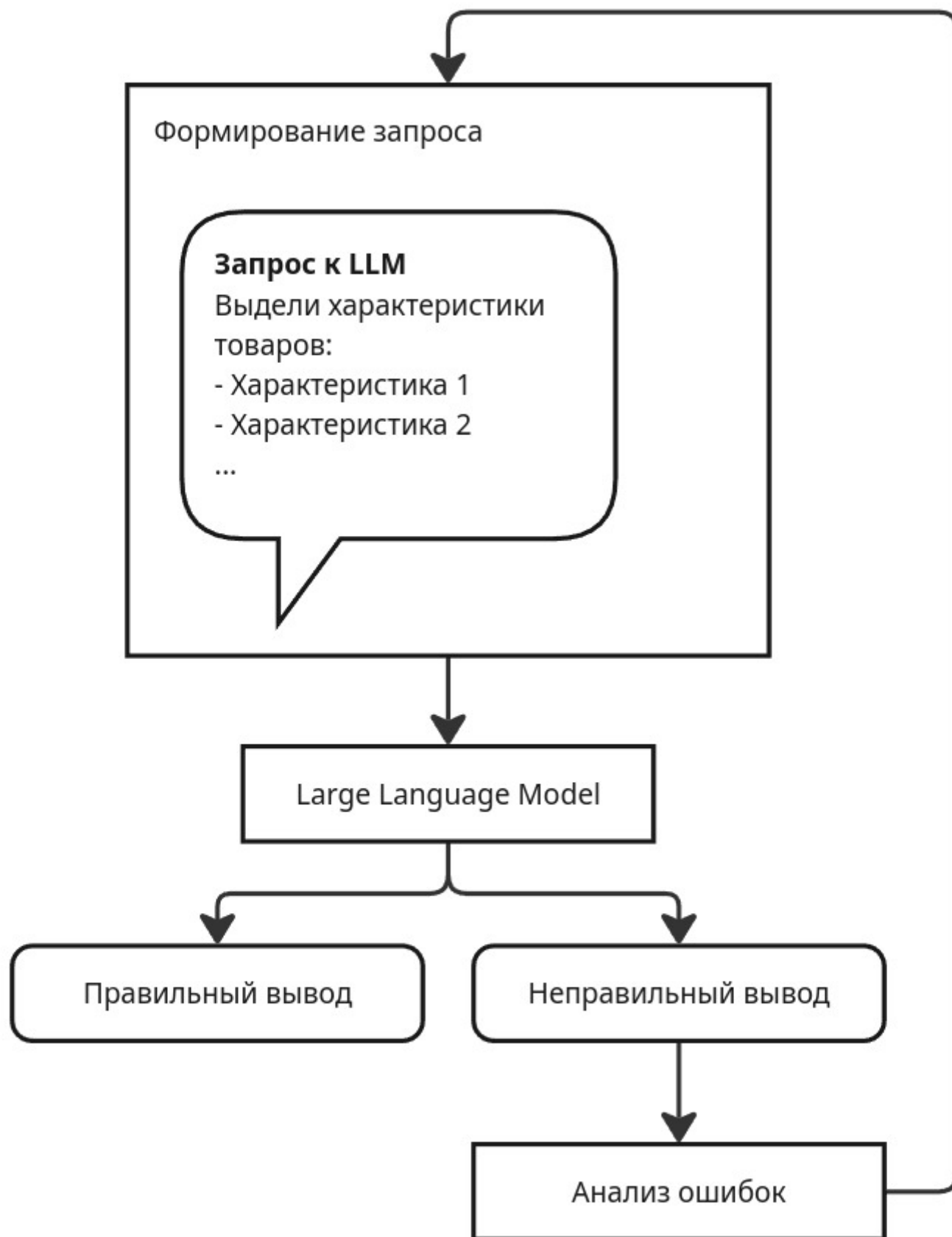
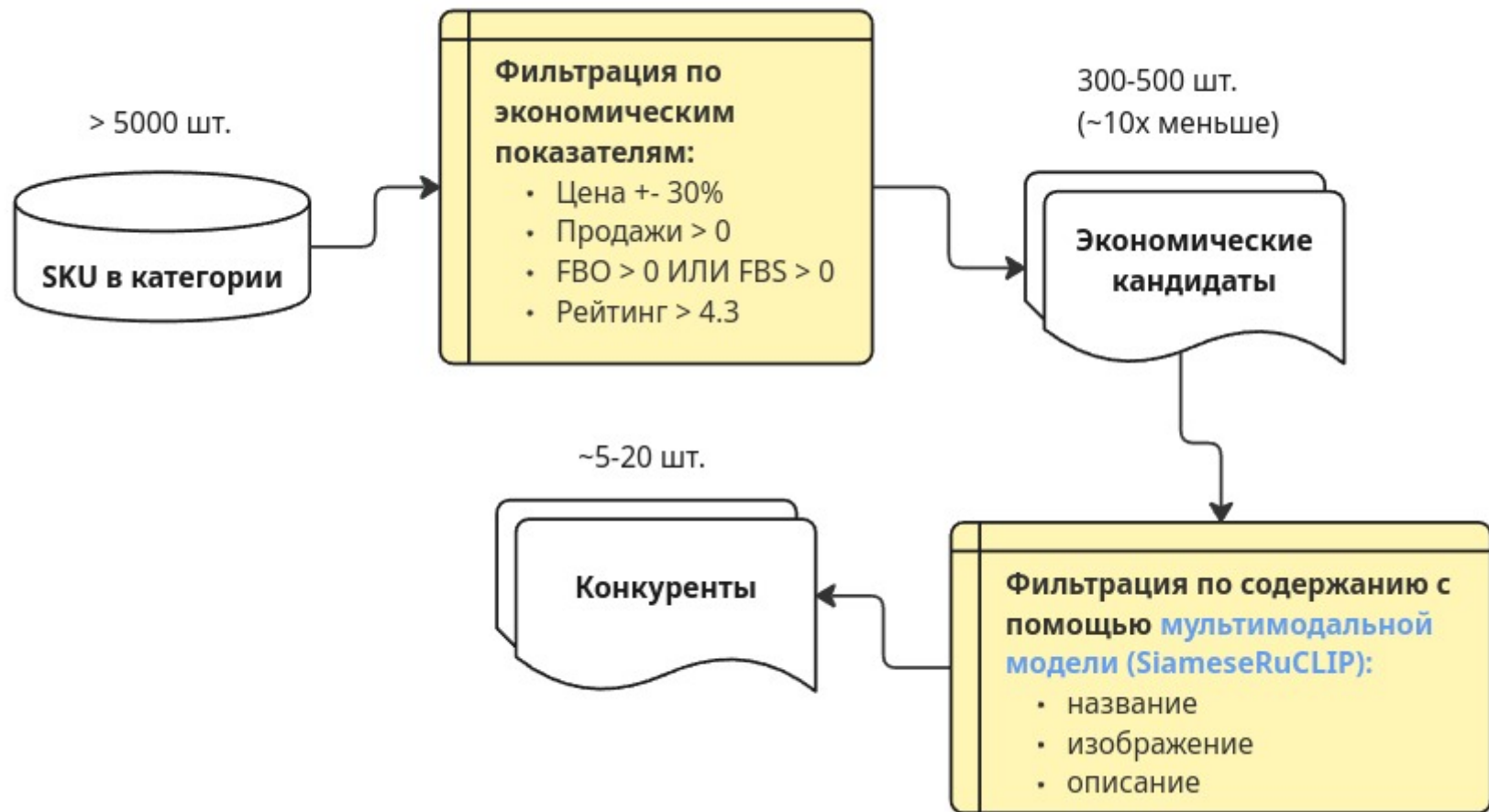


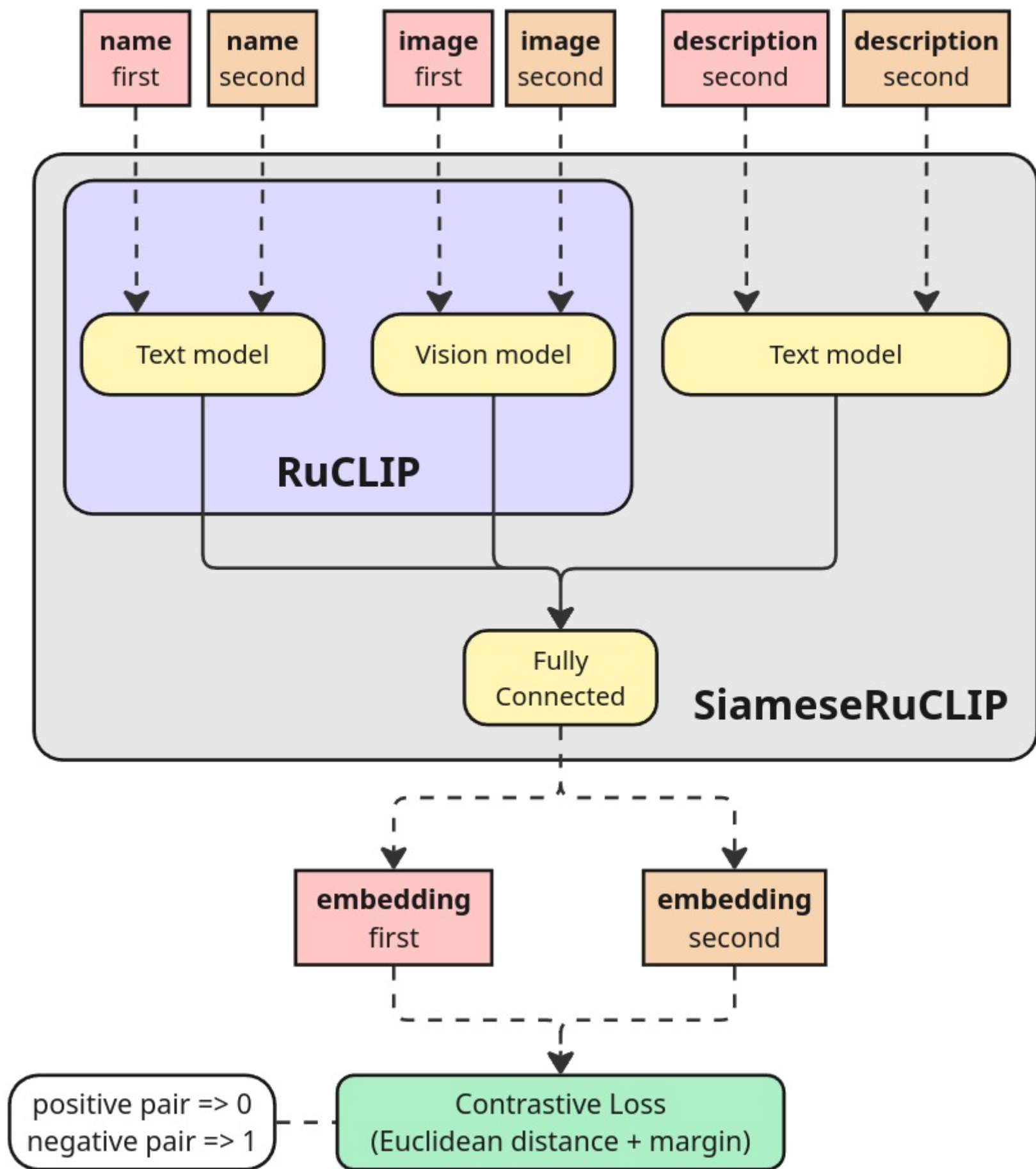
Checkpoint loading



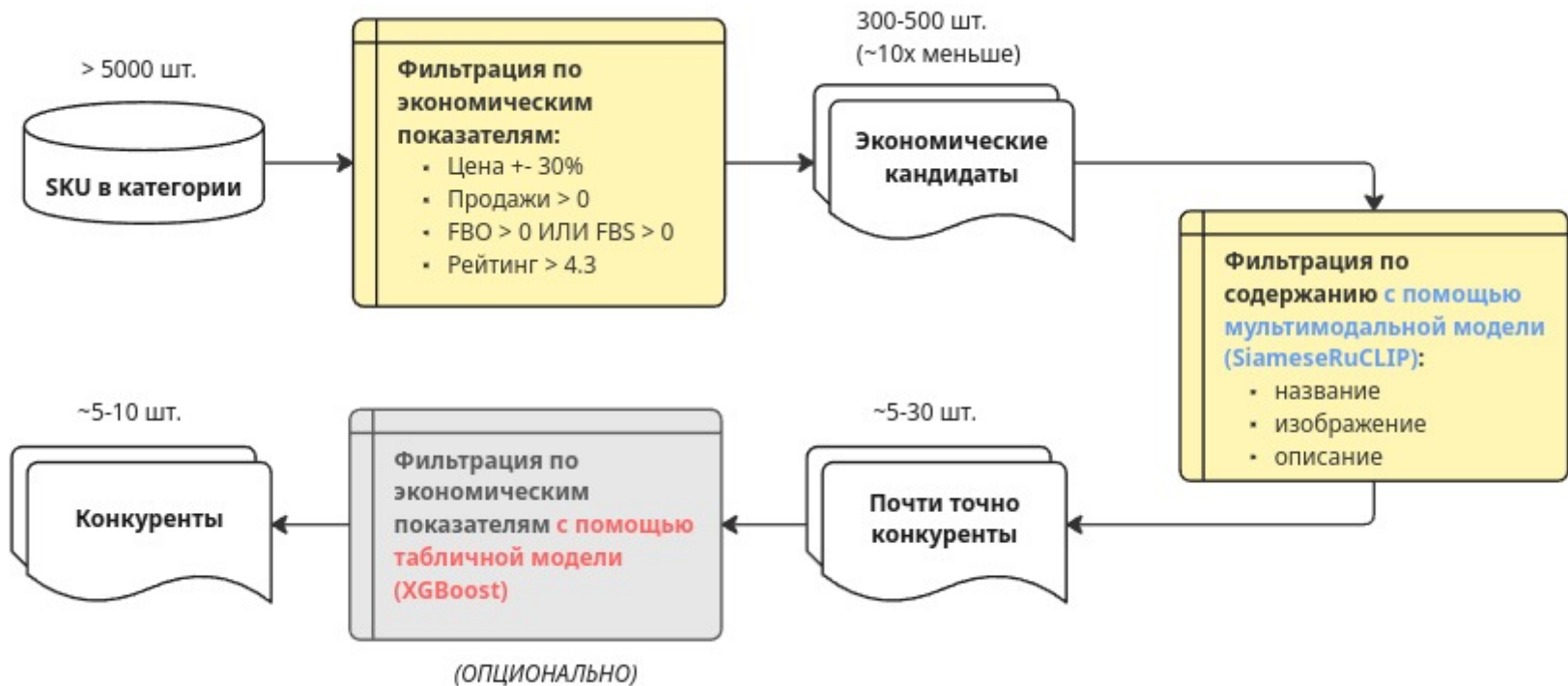








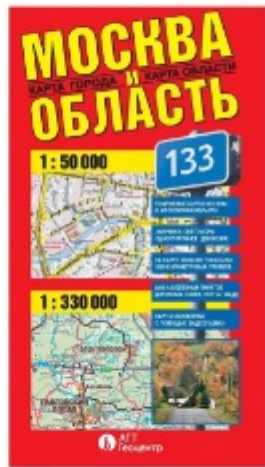




Целевые SKU



Положительные кандидаты



Не отличаются ключевыми характеристиками

Сложные отрицательные кандидаты



Совпадают почти по всем ключевым характеристикам

Простые отрицательные кандидаты



Имеют противоположные характеристики

~5-20 **samples**
(25-400 *pairs*)

positives

~50-100 **samples**
(2,500-10,000 *pairs*)

hard_negatives

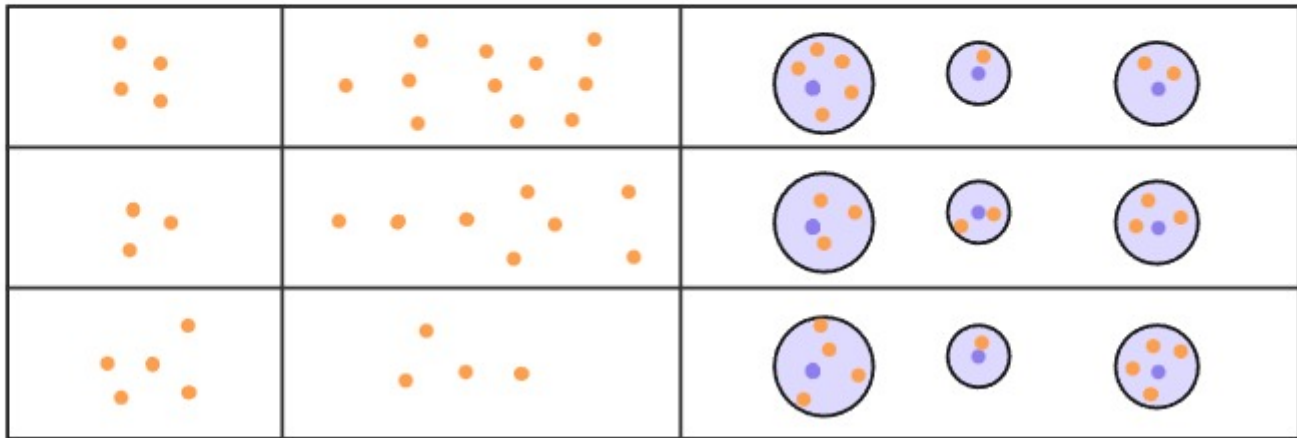
5,000 **samples** -> 500 cluster **medians**
(25,000,000 *pairs* -> 250,000 *pairs*)

soft_negatives

query_sku_1

query_sku_2

query_sku_3



positives

hard_negatives

soft_negatives

--	--	--

Train split (80%)

positives

hard_negatives

soft_negatives

--	--	--

Validation split (10%)

positives

hard_negatives

soft_negatives

<div>query sku 1 query sku 2 ... query sku n</div>		
--	--	--

Test split (10%)

positives

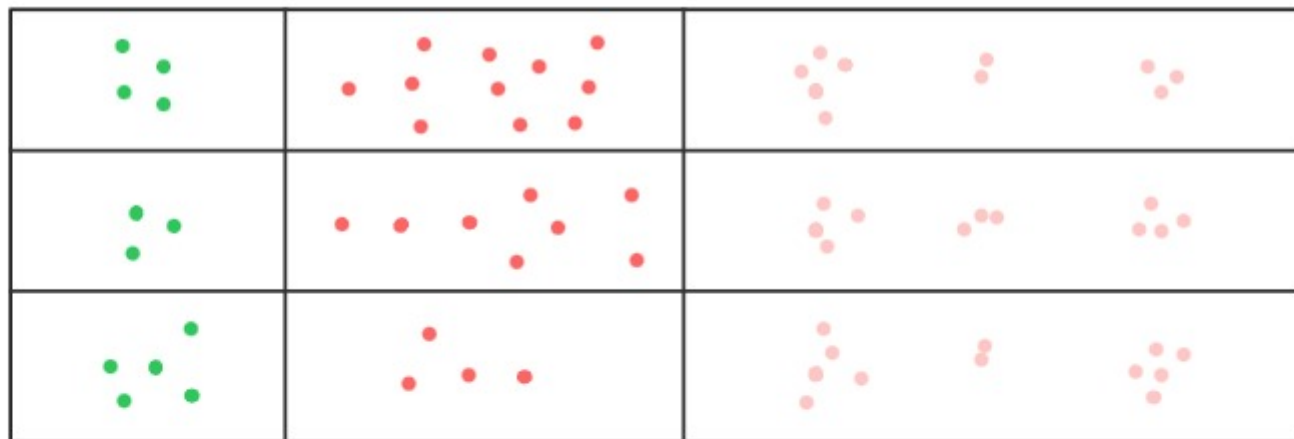
hard_negatives

soft_negatives

query_sku_1

query_sku_2

query_sku_3



positive pairs

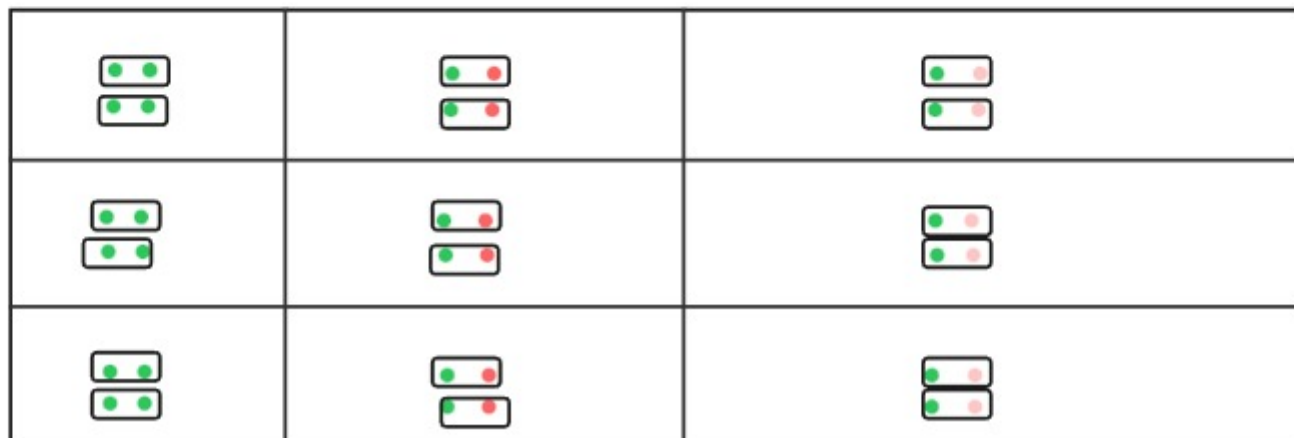
hard neg pairs

soft neg pairs

query_sku_1

query_sku_2

query_sku_3



PAIRWISE_TABLE

sku_first

sku_second

label

query_sku_1

pos_1

1

query_sku_1

hard_neg_1

0

query_sku_1

soft_neg_1

0

query_sku_1

soft_neg_1

0

query_sku_2

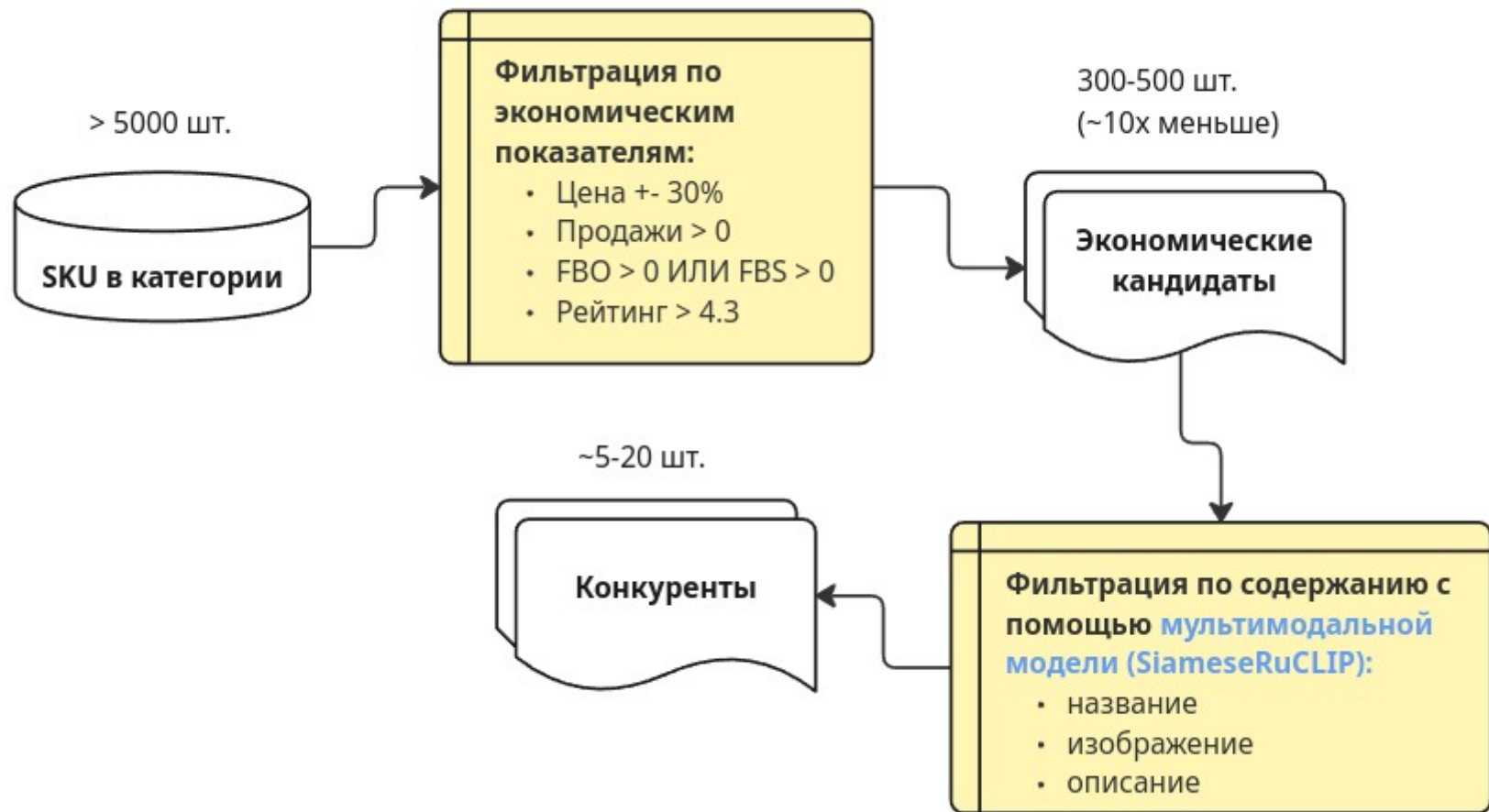
pos_1

1

query_sku_3

hard_neg_1

0



positives

hard_negatives

soft_negatives

query_sku_1



query_sku_2



query_sku_3



positives

hard_negatives

soft_negatives

query_sku_1



query_sku_2



query_sku_3



Train split (80%)

positives

hard_negatives

soft_negatives

query_sku_4



query_sku_5



query_sku_6



Validation split (10%)

positives

hard_negatives

soft_negatives

query_sku_7



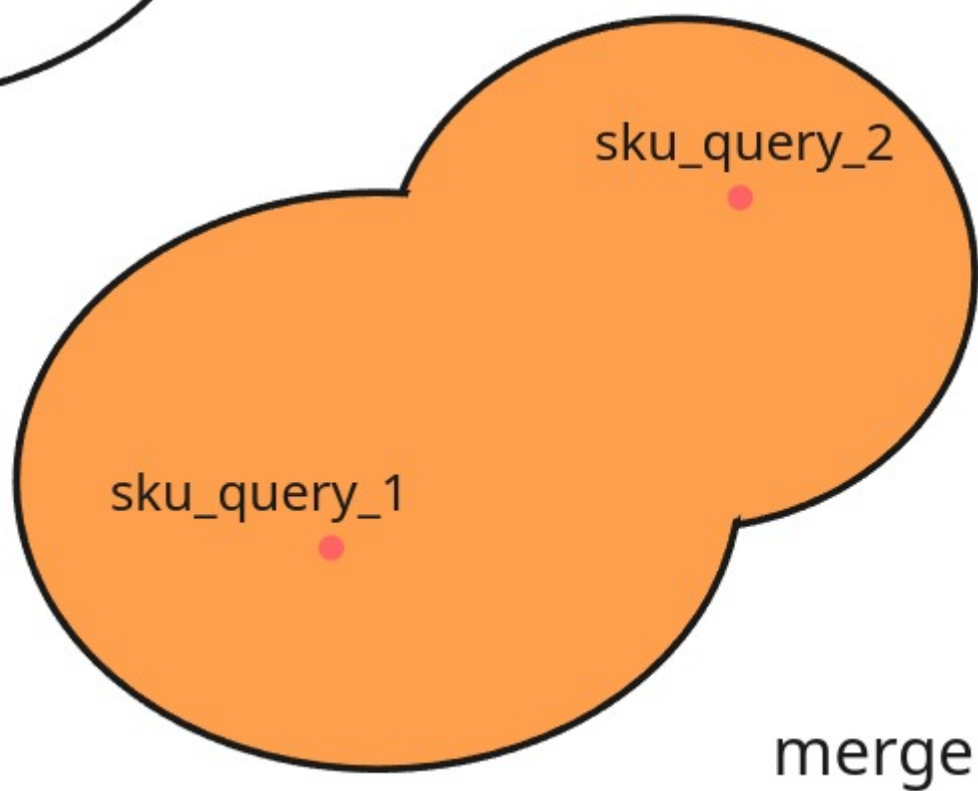
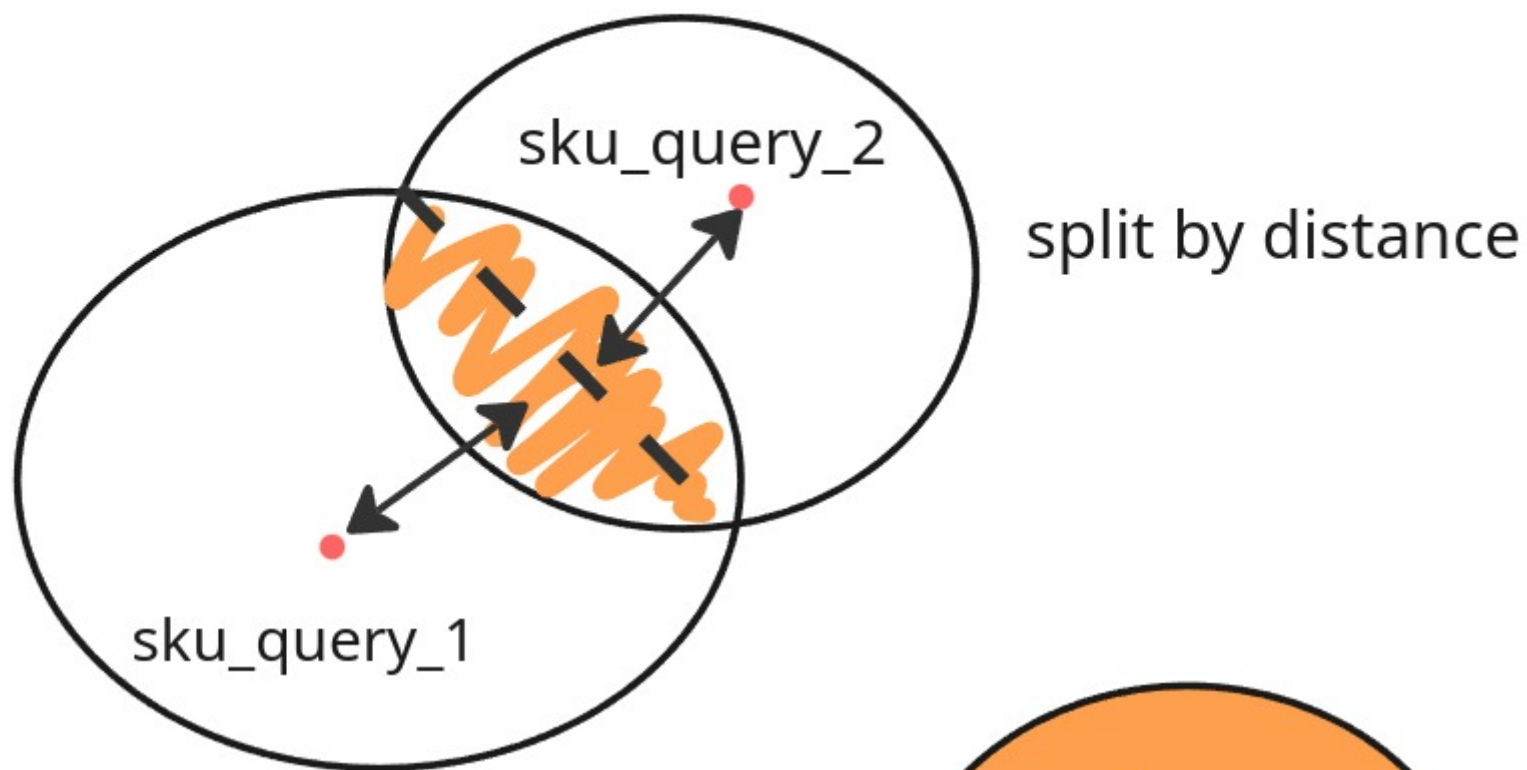
query_sku_8



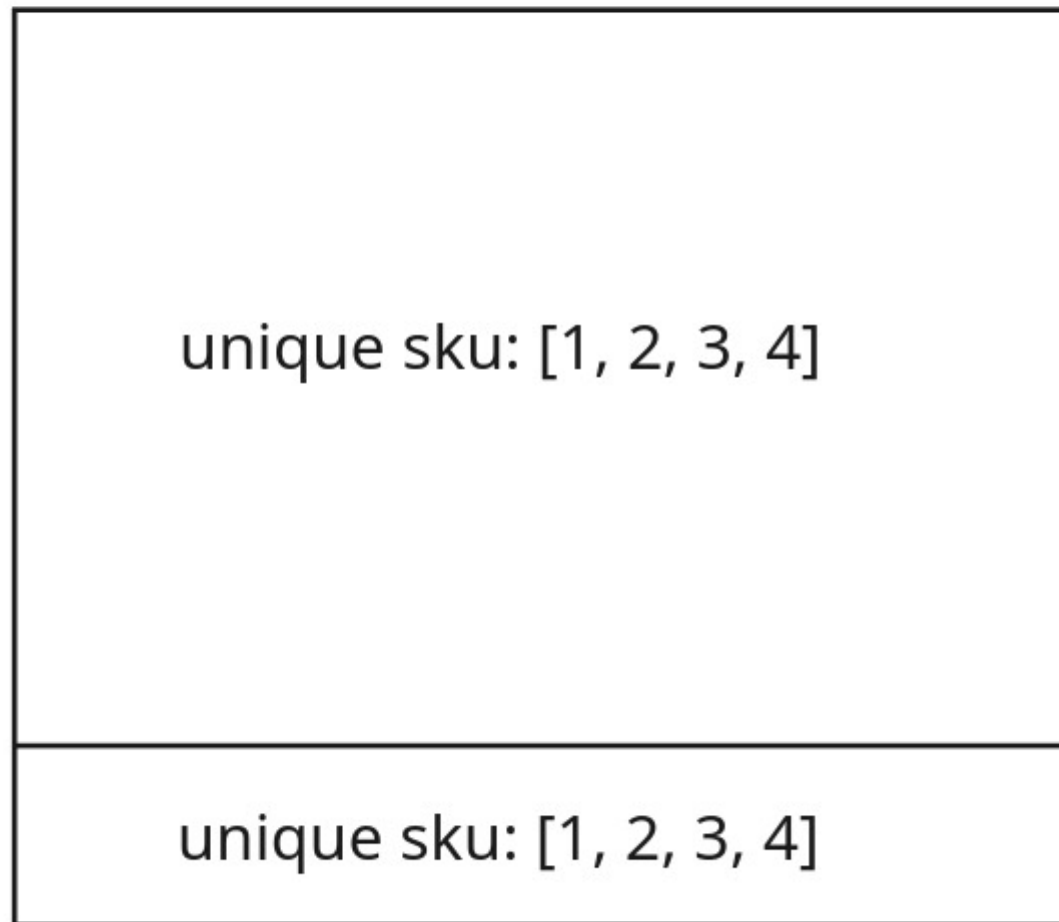
query_sku_9



Test split (10%)



Отсутствие пересечения по уникальным товарам не гарантировано!



Train

Validation

Положительные для одного товара могут быть отрицательными для другого в другого, **в том числе в другом сплите**

	positive pairs	hard neg pairs	soft neg pairs	
query_sku_1	  1, 2	  1, 3	  1, 4	Train
query_sku_2	  3, 4	  4, 2	  3, 1	Validation

positives

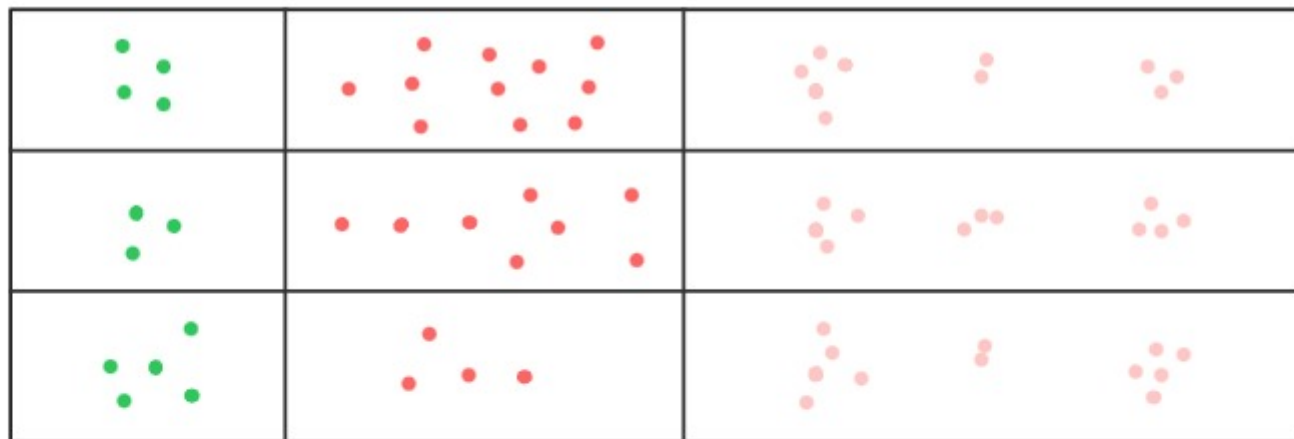
hard_negatives

soft_negatives

query_sku_1

query_sku_2

query_sku_3



positive pairs

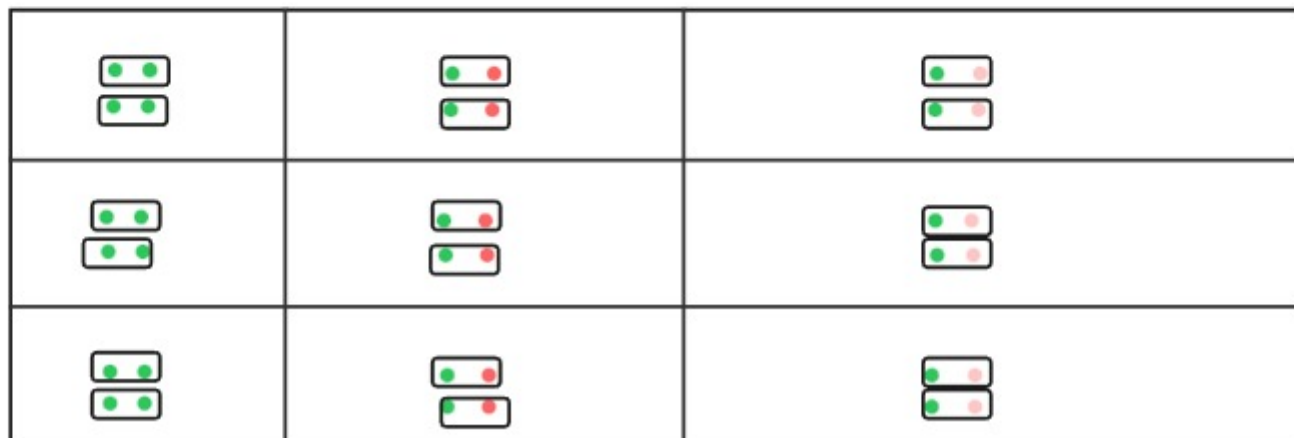
hard neg pairs

soft neg pairs

query_sku_1

query_sku_2

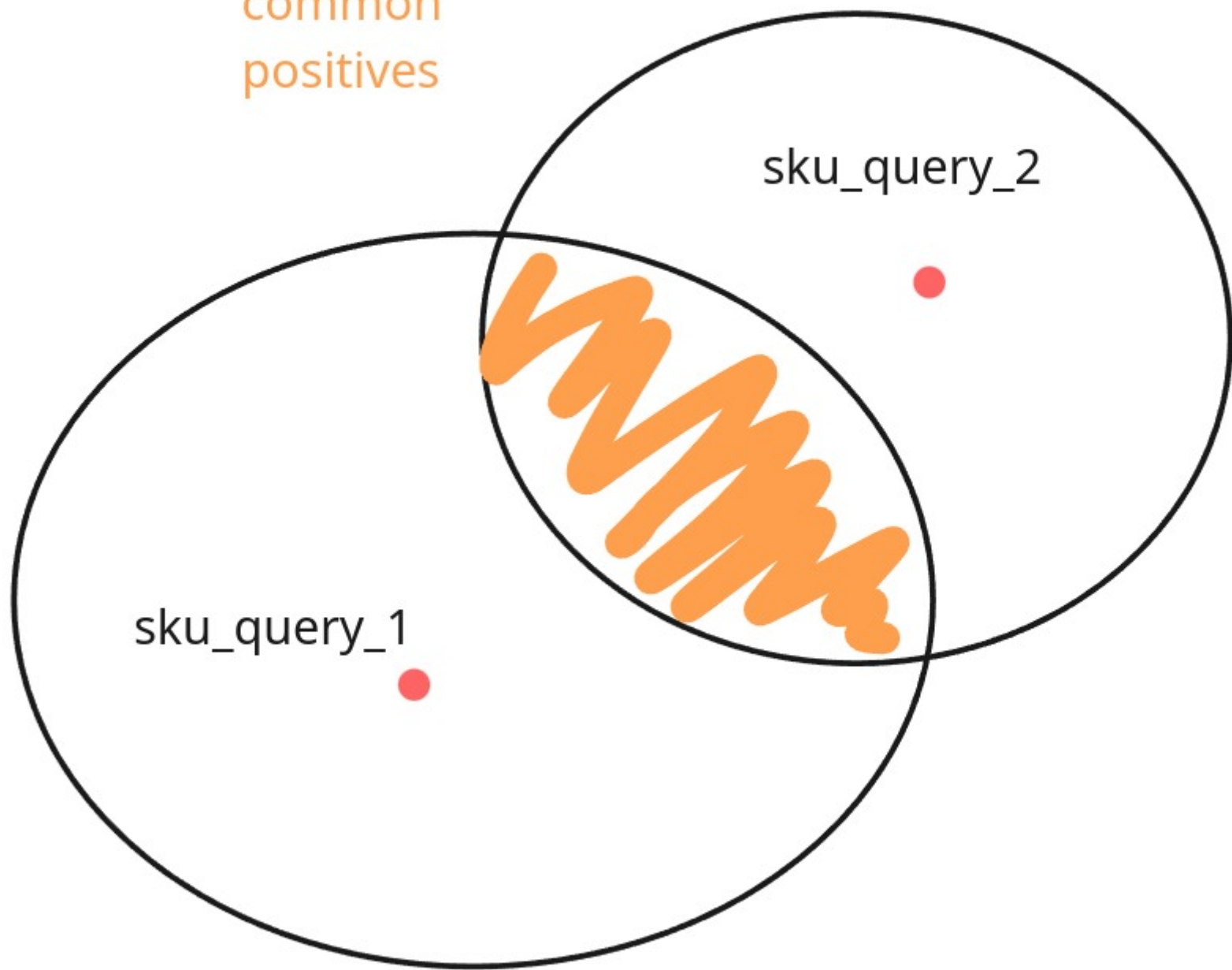
query_sku_3




common
positives

sku_query_2

sku_query_1



Положительные для одного товара могут быть *положительными* для другого в другого, **в том числе в другом сплите**

	positive pairs	hard neg pairs	soft neg pairs	
query_sku_1	 1, 2			Train
query_sku_2	 3, 1			
				Validation