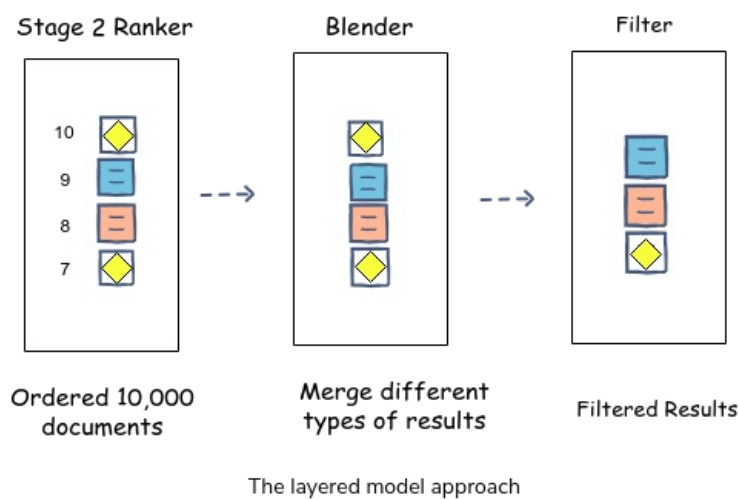


Filtering Results

Let's further filter the results based on how they might be perceived by the searcher despite good user engagement.

We'll cover the following ^

- Result set after ranking
- ML problem
 - Training data
 - Features
 - Building a classifier



Until now, you have selected relevant results for the searcher’s query and placed them in order of relevance. The job seems to be pretty much completed. However, you may still have to filter out results that might seem relevant for the query but are inappropriate to show.

Result set after ranking

The result set might contain results that:

- are offensive
- cause misinformation
- are trying to spread hatred
- are not appropriate for children
- are inconsiderate towards a particular group

These results are **inappropriate despite having good user engagement**.

How do we solve this problem? How do we make sure that a search engine can be safely used by users of all age groups and doesn’t spread misinformation and hatred?

ML problem

From a machine learning point of view, we would want to have a specialized model that removes inappropriate results from our ranked result set.

As we discussed for our main ranking problem, we would need training data, features, and a trained classifier for filtering these results.

Training data

Let's go over a couple of methods that we can use to generate training data for filtering undesired results.

Human raters

Human raters can identify content that needs to be filtered. We can collect data from raters about the above-mentioned cases of misinformation, hatred, etc. and from their feedback, we can train a classifier that predicts the probability that a particular document is inappropriate to show on SERP.

Online user feedback

Nowadays, good websites provide users with the option to report a result in case it is inappropriate. Therefore, another way to generate data is through this kind of online user feedback. This data can then be used to train another model to filter such results.

Features

We can use the same features for this model that we have used for training our ranker, e.g., document word embeddings or raw terms can help us identify the type of content on the document.

There are maybe a few particular features that we might want to add specifically for our filtering model. For example, *website historical report rate*, *sexually explicit terms used*, domain name, website description, images used on the website, etc.

Building a classifier

Once you have built the training data with the right features, you can utilize classification algorithms like logistic regression, MART(Boosted trees or Random Forest), or a Deep neural network to classify a result as inappropriate.

Similar to the discussion in the ranking section, your choice of the modelling algorithm will depend on:

- how much data you have
- capacity requirements
- experiments to see how much gain in reducing bad content do we see with that modelling technique.


← Back


Ranking

Next →

Problem Statement

 Mark as Completed

 Report an Issue

 Ask a Question

(https://discuss.educative.io/tag/filtering-results__search-ranking__grokking-the-machine-learning-interview)