

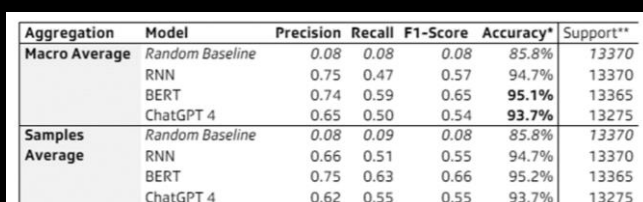
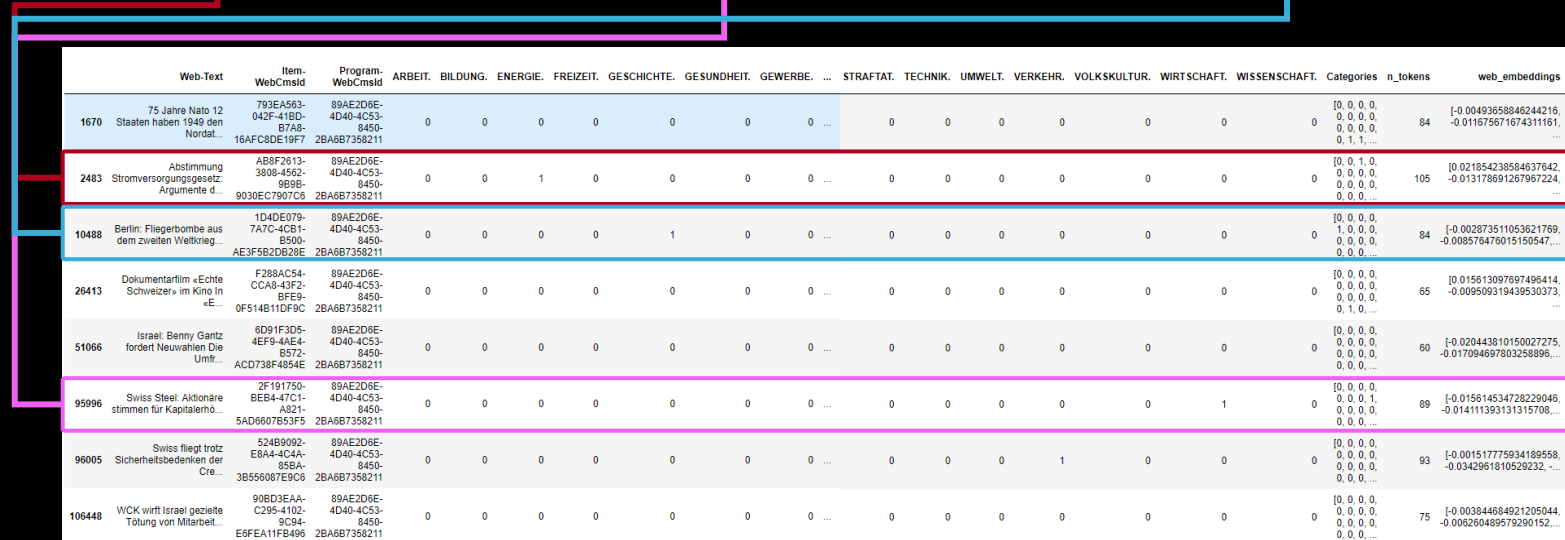
The screenshot shows the SRF news website. At the top, there's a navigation bar with a search icon and the URL <https://www.srf.ch/play/tv/tagesschau>. The main content area features a live broadcast of Cornelia Boesch, the news anchor, with the 'SRF news' logo and a large red globe graphic in the background. Below the main video, there's a row of eight video thumbnails, each with a title, a small image, and a duration. The thumbnails are: '75 Jahre Nato' (5:46), 'Israel: Benny Gantz fordert Neuwahlen' (1:44), 'WCK wirft Israel gezielte Tötung von Mitarbeitenden vor' (3:03), 'Swiss fliegt trotz Sicherheitsbedenken der Crews nach Tel Aviv' (2:33), 'Abstimmung Stromversorgungsgesetz: Argumente der...' (2:34), 'Swiss Steel: Aktionäre stimmen für Kapitalerhöhung' (2:09), 'Berlin: Fliegerbombe aus dem zweiten Weltkrieg ist entschärft' (2:45), and 'Dokumentarfilm «Echte Schweizer» im Kino' (2:54). A red line connects the 'Swiss fliegt...' thumbnail to the 'Swiss Steel' thumbnail, and a pink line connects the 'Swiss Steel' thumbnail to the 'Berlin: Fliegerbombe...' thumbnail. A blue line connects the 'Berlin: Fliegerbombe...' thumbnail to the 'Dokumentarfilm...' thumbnail.



UT  
04.04.2024 · 3 Min

UT  
04.04.2024 · 2 Min

UT  
04.04.2024 · 3 Min



- ❖ Macro Average: Calculates metrics individually for each label and then averages them without considering label imbalance, resulting in an unweighted mean. This is perfectly fine because our dataset was as balanced as possible.
- ❖ Samples Average: Evaluates the metric for each sample based on its true and false predicted classes and returns the average of these sample-level metrics.
- ❖ Precision: Correct positive predictions over total predicted positives.
- ❖ Recall: Correct positive predictions over actual positives.
- ❖ F1 Score: Harmonic mean of Precision and Recall.

\*Hamming Accuracy: Fraction of exact matches in multi-label classification. Hamming Accuracy was computed for Macro and Sample Average. The random baseline model shows that by guessing randomly the Accuracy would still be 85.8% or higher due to the huge number of true negatives in the dataset.

\*\* The Support refers to the overall occurrence of positive labels in the test set. There are minimal differences between the models because of the different max-token length limitations between ChatGPT 4, the OpenAI Embedding API and BERT.