

**Online Appendix A: Definitions of Relevance in the Literature**

Table 1

*Overview of Various Definition of Comment Relevance Used in the Literature*

Reference	Definition of relevance used
The present paper	A comment is relevant if it reflects individuals' experiences of, or opinions or thoughts about (the content of) a video, or the artist/maker of the video. A comment is also relevant if it relates to the social experience of watching a video on a social media platform.
Shetty, A., Abreo, B., D'Souza, A., Kondana, A., & Kavitha, K. M. (2021). Video description based Youtube comment classification. <i>Proceedings of First Global Conference on Artificial Intelligence and Applications (GCAIA 2020)</i> , 667-678. <a href="https://doi.org/10.1007/978-981-33-4604-8_51">https://doi.org/10.1007/978-981-33-4604-8_51</a>	A comment is relevant if its content is closely related or appropriate to what is discussed in the video, whereby comments containing a positive or negative video evaluation are regarded as a distinct category of comments.
Dubovi, I., & Tabak, I. (2020). An empirical analysis of knowledge co-construction in YouTube comments. <i>Computers &amp; Education</i> , 156, 1–16. <a href="https://doi.org/10.1016/j.compedu.2020.103939">https://doi.org/10.1016/j.compedu.2020.103939</a>	A comment is relevant if it leads to the co-construction of knowledge related to a video.

Poché, E., Jha, N., Williams, G., Staten, J., Vesper, M., & Mahmoud, A. (2017). Analyzing user comments on YouTube coding tutorial videos. *2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC)*, 196–206. <https://doi.org/10.1109/ICPC.2017.26>

Zweig, L., Liu, C., Hiraga, M., Reed, A., Czerniakowski, M., Dickinson, M., & Kübler, S. (2017). FunTube: Annotating funniness in YouTube comments. *Proceedings of the Workshop on Corpora in the Digital Humanities (CDH)*, 48-57. Retrieved from: <https://ceur-ws.org/Vol-1786/zweig.pdf>

Serbanoiu, A., & Rebedea, T. (2013). Relevance-based ranking of video comments on YouTube. *2013 19th International Conference on Control Systems and Computer Science*, 225–231.

A comment is relevant if it is useful for the creator of a video because it includes questions or concerns about the video content, points out errors in the video, requests the production of additional future content, or makes suggestions to improve the quality of the video.

A comment is only marked as relevant if it is relevant to the content of a video. The content of a video includes information about the title, actors, specific jokes, or dialogue in the video. It also includes general reactions to the content of a video, even when this content is not directly discussed. It does not include aspects of the video other than its content, such as the cinematography, lighting, music, setting, or general topic of the video. Comments that are not clearly understandable (e.g., because they are not written in English) are also not considered as relevant comments.

Relevance is determined based on the following textual features: (a) the number of non-ASCII-characters, (b) the number of capital letters, (c) the number of new-lines, (d)

<https://doi.org/10.1109/CSCS.2013.87>

the number of digits, (e) the number of trivialities, (f) the number of words, (g) the mean word length, (h) the number of punctuation marks, and (i) a predefined list of frequent spam text patterns.

Ammari, A., Dimitrova, V., & Despotakis, D. (2012). Identifying relevant YouTube comments to derive socially augmented user models: A semantically enriched machine learning approach. *In Advances in User Modeling: UMAP 2011 Workshops, Spain*, 71-85. [https://doi.org/10.1007/978-3-642-28509-7\\_8](https://doi.org/10.1007/978-3-642-28509-7_8)

Focusing on comments written in response to videos about job interviews, the authors indicate that a comment is relevant if it well describes the job interview domain or if the comment can be written by a user who wants to learn about the job interview domain or reflects characteristics of such comments.

---

Please note that this table is not meant to provide a complete overview of all the published work on the relevance of online comments, but merely as an indication of the variety in the definition of relevance in the literature.

### Online Appendix B: Cross Validation Results

Table 1

*Means and Standard Deviations of  $F_1$ -scores Obtained for the Best Performing Models Through 5-fold Cross Validation*

---

Classifier	$M_{F1-score}$	$SD_{F1-score}$
------------	----------------	-----------------

---

English (LR, Count vectorizer)	0.87	0.00
Relevance (LR, Count vectorizer)	0.82	0.01
Positive evaluation (LR, Count vectorizer)	0.80	0.00
Negative evaluation (C4.5 decision tree, Count vectorizer)	0.97	0.00
Providing information (C4.5 decision tree, Count vectorizer)	0.91	0.00
Requesting information (C4.5 decision tree, Count vectorizer)	0.99	0.00
Personal experience (LR, Count vectorizer)	0.86	0.00
Community (LR, Count vectorizer)	0.86	0.01
Other (C4.5 decision tree, Count vectorizer)	0.92	0.00

---

*Table 1 includes the following abbreviation: LR (Logistic Regression).*