Journal of the Brazilian Computer Society, 2022, 28:1, doi: 10.5753/jbcs.2022.2216

Using Visual Features and Early Views to Classify the Popularity of Facebook Videos

Bruna Martini Dalmoro [**Pontifical Catholic University of Rio Grande do Sul** | ***bruna.dalmoro@edu.pucrs.br***]
Soraia Raupp Musse [**Pontifical Catholic University of Rio Grande do Sul** | ***soraia.musse@pucrs.br***]

Pontifical Catholic University of Rio Grande do Sul, Av. Ipiranga, 6681, Partenon, Porto Alegre, RS, 90619-900, Brazil.

Received: 15 September 2021 • **Accepted:** 24 October 2022 • **Published:** 28 December 2022

Abstract These days it is easy to create and share content online. Millions of people create and share their online content, that is consumed by millions more, daily. This flow of content and consumption has been used as a channel for disseminating digital advertisements, generating publicity for brands and financial return for content creators. Thus, identifying whether a video will be popular in the first moments after its publication is of great value to advertisers. Using Random Forest, we classify Facebook videos as popular or unpopular based on their number of views, using early views and visual features extracted from the videos as predictor features. Our results indicate that using the combination of early views with visual features yields the best results, allowing the prediction of popularity to be made as early as possible.

Keywords: Popularity, Video Analysis, Visual Features, Random Forest

1 Introduction

Social networks are driven by the posts of their users, who create and share content. Videos are popular formats for sharing content, whether via YouTube, Facebook, Instagram or TikTok. According to numbers released on Youtube for Press [2022], over 500 hours of videos are published per minute, accessed by 2 billion monthly users who generate billions of views daily. Regarding content creators, the number of channels with more than one million subscribers grew by more than 65% per year. When it comes to revenue, the number of channels that had a six-digit annual revenue on YouTube grew more than 40% per year.

According to the site "About Facebook [2018]", right after the launch of Facebook Watch (Facebook's video viewing tool), over 400 million people a month and 75 million per day spend an average of more than 20 minutes on the Watch. Two years later, according to the same source (About Facebook [2020]), over 1.25 billion people visit the Watch every month to discover and share videos from millions of creators and publishers. This flow of users and content can be interesting for companies that want to promote their content online and reach a large audience. Therefore, understanding what makes a video popular and being able to predict its popularity is a problem that companies like Facebook and Netflix have invested in solving. This predictive power is useful both for advertisements, since they can be directed to videos of greater reach, and for content creators, regarding the management and production of content based on characteristics that generate more views.

In our previous work Dalmoro and Musse [2021], our goal was to use only visual features taken from Facebook videos to predict their popularity using Support Vector Machine classifier. In this work, we use the same dataset now including other feature as the number of early views in addition to the visual features extracted from 1,820 videos posted on Facebook to predict the popularity of the videos, as described in Section 3. We use a Random Forest classifier, developed with the caret package in the R software, and compare the results of our classification model with the regression model provided in previous competitive work by Trzciński and Rokita [2017]. Details on the method are presented in Section 4. Our results, presented in Section 5, show that the selected visual features, presented in the paper, are related to the number of views of a video, improving the results when combined with the number of early views. We've also shown that Random Forest sorting is effective for these types of tasks.

2 Related Work

The sharing and consumption of content on social networks has been the reality of a large part of the population, and it has been seen as an opportunity to reach a greater audience with digital advertising. Thus, one of the challenges is to identify the most relevant content that can give more visibility to the ad served. The prediction of online video popularity is a problem already explored in the literature in several ways. Zohourian *et al*. [2018] used Instagram image and video context features to predict the number of likes of a post through regression and classification methods. Ouyang *et al*. [2016] built regression models to predict future view count values and validated their approach on a video dataset from China's online video service Youku.

Trzciński and Rokita [2017] used Support Vector Regression with Radial Base Function with Gaussian kernel. Using data collected from Facebook pages, the authors propose a method called Popularity-SVR, that predicts popularity of an online video using Support Vector Regression (SVR). The

Facebook video data included visual features and temporal features, that is, features captured soon after the content was published, such as number of views over time. To assess the performance of the proposed predictive model, they used Spearman's correlation, proposed by Spearman [1904], a non-parametric measure of statistical dependence between two variables. This measure ranges from -1 to 1, where -1 indicates a perfect inverse relationship between these variables, 1 indicates a perfect positive relationship and when the relationship between them is closer to 0, the relationship between them is smaller. When it comes to visual features, the Popularity-SVR shows that, individually, deep features provide the highest Spearman's correlation value with video popularity (0.13), followed by the feature groups Clutter (0.12) and Scene Dynamics (0.08). Overall correlation value using all visual features reached over 0.23. However, better results were obtained when visual features were combined with the number of views, where the Spearman's correlation reached over 0.93.

In our previous work (Dalmoro and Musse [2021]), our goal was to use only visuals taken from Facebook videos to predict their popularity. We used Support Vector Machine with Radial Gaussian Basis Function to classify these videos into two groups: the most and the least popular, according to the number of views. Our best model obtained a Kappa of 0.7324, sensitivity of 0.8930 and positive predictive value of 0.8930, using video characteristics and rigidity information. In the present work, we compared the results of a second classification technique, Random Forest, and included the information on the number of early views in the first 1 hour, 6 hours, 12 hours and 24 hours.

3 Dataset and Features

The dataset[1] used in this work was available by Trzciński and Rokita [2017]. The available file contains features extracted from 1,820 videos published on Facebook between August 1st and October 15th 2015 from pages such as AJ+[2] and BuzzFeedVideo[3]. Two types of features were considered in the data extract: temporal and visual features. The next sections describe some details about both features data.

3.1 Temporal Features

Extracted after the video is posted, temporal features contain the number of views, likes, comments and shares in every hour for seven days after posting, collected by the URL scraper on the posting page. In Dalmoro and Musse [2021], we used only visuals as predictor variables, while temporal features were excluded from the analysis, with the exception of the number of views at the end of the period, which was used as a response variable. In the present work, in addition to using the number of views as the response variable, we also used the number of early views, i.e. views recorded **1**, **6**, **12** and **24 hours** after the video was published, as predictor variables to improve prediction results soon in the first moments after publication.

3.2 Visual Features

The visual features were collected directly from the video, using various computer vision algorithms, as described in Trzciński and Rokita [2017]. The list of available visual data that was used in this work is:

1. **Video characteristics:** This class regards general video information, such as duration, frames per second, number of frames, and frame dimensions of the analyzed video.
2. **Dominant color:** The color space of the video was divided into 10 classes (black, white, blue, cyan, green, yellow, orange, red, magenta, and other) and each frame of each video was assigned to one of these classes. In addition, the data set contains information about which class of colors is dominant and what proportion of each color is present for each video.
3. **Face detection:** Information about the presence of faces in the video, such as the average number of faces per frame, the proportion of frames with faces, and the average proportion of the face size in relation to the size of the frame, using a face detector based on a cascade classifier.
4. **Text detection:** Similar to face detection, it refers to the presence of text in the video, such as the proportion of frames with text and the average proportion of the text size in relation to the size of the frame, using a combination of edge detection, morphological filters and Tesseract-OCR engine.
5. **Scene dynamics:** It regards information about the number of shots in the video and classification of the shots as hard and soft cuts, using Edge Change Ration algorithm.
6. **Rigidity:** They provide information about the average video speed, a clutter metric, and a metric that specifies the video rigidity, using a combination of FAST feature point detector and BRIEF descriptor.

More details about the features can be found in Trzciński and Rokita [2017] work.

While Trzciński and Rokita [2017] propose the popularity-SVR, treating the popularity prediction as a regression problem, in our work, we treat it as a classification problem. We test two different ranking techniques and try to predict whether or not a video will be popular based on different milestones. We have tested five different milestones that we think might be interesting, as described below.

4 Method

In this work, our goal is to predict whether or not a video will be popular based on the number of views 7 days after publication, given its visual features computed using computer vision algorithms, and to observe the improvement in classification performance including the number of views in the first 1, 6, 12 and 24 hours after publication. This section presents the pre-processing phase performed on the available

[1] http://ii.pw.edu.pl/~ttrzcins/facebook_dataset_2015.csv
[2] https://www.facebook.com/ajplusenglish
[3] https://www.facebook.com/BuzzFeedVideo

Table 1. Combinations of group of features, as defined in Section 3.2, tested in the models and abbreviations that we use to refer to the feature setup.

Abbreviation	Combination of Visual Features
V	Video Characteristics
C	Dominant Color
F	Face Detection
T	Text Detection
D	Scene Dynamics
R	Rigidity
VC	V. Char. + Color
VF	V. Char. + Faces
VT	V. Char. + Text
VD	V. Char. + S. Dyn.
VR	V. Char. + Rigidity
VDC	V. Char. + S. Dyn. + Color
VDF	V. Char. + S. Dyn. + Faces
VDT	V. Char. + S. Dyn. + Text
VDR	V. Char. + S. Dyn. + Rigidity
VDRC	V. Char. + S. Dyn. + Rigidity + Color
VDRF	V. Char. + S. Dyn. + Rigidity + Faces
VDRT	V. Char. + S. Dyn. + Rigidity + Text
VDRTC	V. Char. + S. Dyn. + Rigidity + Text + Color
VDRTF	V. Char. + S. Dyn. + Rigidity + Text + Faces
Complete Model	V. Char. + S. Dyn. + Rigidity + Text + Color + Faces

data, the model tuning to configure the hyperparameters, and details about the Random Forest classifier model used. All analyzes and modeling were performed using the R software version 3.6.3, maintained by R Core Team [2020], through the caret package version 6.0-86, by Kuhn *et al*. [2008].

As in Dalmoro and Musse [2021], we treated popularity prediction as a classification problem. In this case, we do not want to provide the exact number of views for a given video, but rather identify whether the video will have more views than a certain pre-established milestone 7 days after its publication. We tested 5 different milestones according to the number of views: **10,000**, **100,000**, **500,000**, **750,000** and **1 million views**. We call the videos that reached the milestone as *successful-videos*. More details are presented in Section 5.

In addition, we compared the results of models that use only visual features as predictor variables with models that combine the visual features with view number information in the first few hours after publication. We tested using the first **1**, **6**, **12** and **24 hour views**.

4.1 Pre-processing

According to Kuhn and Johnson [2013], data preparation can make or break the predictive power of a model. As in Dalmoro and Musse [2021], in pre-processing we seek to identify and correct missing values, zero- and near zero-variance predictor features, linear correlations and dependencies.

We identified four correlated feature predictors: number of frames and video duration, frame width and frame height, average proportion of frames with faces and average number of faces per frame and two features about soft and hard cuts are complementary, so they have a correlation of -1. Consequently, the features on number of frames, frame width, average proportion of frames with faces, and one of the two features on shot clipping have been removed.

The training and testing sets are randomly divided for each model in proportions of 70% and 30%, respectively, preserving the general distribution of the response variable class. Finally, we centralized and scaled the data to improve the numerical stability of some calculations, as indicated by Kuhn and Johnson [2013].

4.2 Random Forest Classifier

The technique used for predictive modeling of video popularity is Random Forest. Proposed by Breiman [2001], Random Forest is a popular decision tree-based supervised learning algorithm. Decision trees classify the observations through a rule built by recursive binary partitioning, with the objective of separating as much as possible the classes of the predictor feature in different nodes. At each step of fitting a classification tree, an optimization is performed to select a node, predictor feature, and cut that result in the most homogeneous subgroups for the data. Random Forest fits many classification trees to a dataset and then combines the predictions from all the trees. The algorithm starts with selecting several bootstrap samples of the data, and for each one of them adjusts a classification tree. Observations from the original dataset that are not in a bootstrap sample are called out-of-bag observations. Each binary partitioning of each tree is done based on a small number of features. From the development of all trees, each one of them is used to predict the out-of-bag observations, in a similar way to cross validation. The predicted class of an observation is calculated by the majority of the votes of the out-of-bag predictions for that observation, with randomly divided ties. In this work, we use the caret package to perform the analyses, and the hyperparameter to be adjusted is *mtry*, that is, the number of randomly selected variables to divide each node.

4.3 Model Tuning

For each views milestone, we tested 21 different combinations of visual features, according to Table 1. To set the value for the hyperparameter *mtry* that results in the best model, we create a grid of values between 1 and 25. Then, in the model tuning process, we use repeated 10-fold cross-validation, where we performed 3 different 10-fold cross-validations with different samples, to obtain a less biased estimate of the model result as recommended by Rodriguez *et al.* [2010]. For each combination of visual features and milestones, the hyperparameter that generated the model with the largest Kappa was selected, thus obtaining 105 models.

From these 105 models, we selected the best results based on Kappa, sensitivity and positive predictive value metrics, as detailed in Section 5. Using the best combination of visual features and milestone, we retrain the model according to model tuning process defined, including the number of views obtained in the first 1, 6, 12 and 24 hours after publication.

Finally, in order to compare the benefit of using visual features combined with views information in the first few hours after the video is posted, we train models with only views information as predictor features alone. Following the model tuning process defined, we train a model for each combination of predictor feature and milestone, generating 20 models in total.

5 Results

In this section we present the Random Forest classification results for Facebook videos using only visual features, using only early views, and using visual features with early views. Finally, we compare the results obtained with previous work.

5.1 Only Visual Features

In order to predict whether a video will be a *successful-video* based on a certain milestone, we adjusted 21 different models, according to the feature combinations shown in Table 1, for each of the 5 milestones (10,000, 100,000, 500,000, 750,000 and 1 million views), generating a total of 105 adjusted models. To select the best models among the 105, three different metrics were used: Kappa, Sensitivity and Positive Predictive Value.

According to Vieira *et al.* [2010], Cohen's Kappa Coefficient is a statistical measure of agreement between classifications, which compares the model's classification with the response variable more robustly than accuracy since it takes into account the chance of the result being the result of chance. Sensitivity is the ability of a model to identify positive cases, that is, the percentage of *successful-videos* correctly classified in the model among all *successful-videos* in the dataset. While the Positive Predictive Value measures how many true positives are actually positive, that is, how many of the videos are classified as *successful-videos*.

Based on the classification of strength of compliance proposed by Landis and Koch [1977], we considered as the best models those that resulted in Kappa with moderate strength as agreement or more, that is, $Kappa >= 0.41$, in addition to having a sensitivity and positive predictive value of at least 0.5. Based on these metrics, 21 models were selected, shown in Table 2.

As we can see in the results presented in Table 2, just like in Dalmoro and Musse [2021], the milestone that got the best results was 100,000 views. We believe that the balance of the sample in this milestone is related to the better performance of the models, with the other milestones being more unbalanced in the number of *successfull-videos*. It is important to notice that current method using Random Forest, while in a previous work, Dalmoro and Musse [2021], we use Support Vector Machine classifier.

The model highlighted in bold in Table 2 was considered the best model for having the best performance according to the Kappa metric, and for having one of the best performances in terms of Sensitivity and Positive Predictive Value. This model used as a milestone the value of 100,000 views and combines the visual resources **Video Characteristics + Scene Dynamics + Rigidity + Text + Color**, as defined in Section 3.2. The hyperparameter that generated the best model was $mtry = 18$, obtaining Kappa of 0.7276, sensitivity of 0.8991 and positive predictive value of 0.8855.

5.2 Only Early Views

We also train, for each milestone, models using only information from early views, i.e. views obtained at 1, 6, 12 and 14 hours after the video is published, as isolated predictor features. The combination of early views with milestones generated 20 models in total, as shown in Table 3. The best model in terms of Kappa, highlighted in bold, obtained Kappa of 0.9088, sensitivity of 0.9511 and positive predicted value of 0.9749 using the number of views in the first **24 hours** as a predictor variable for the milestone of 100,000 views. Since our goal is to predict popularity based on views seven day after publication, the closer to that time, the easier it is to make this prediction based on the number of views achieved. However, even a few hours after publication, the model provides good results, reaching a Kappa of 0.8702 using views in the initial 6 hours.

5.3 Visual Features and Early Views

To evaluate the use of visual features together with the early views and compare the results, we used the combination of visual features and the milestone that generated the best model, according to Table 2, and we retrained using together the number of views in the first 1, 6, 12 and 24 hours. The results are shown in Table 4. The best model obtained using visual features and number of views as predictors obtained Kappa of 0.9358, sensitivity of 0.9480 and positive predicted value of 1, using as predictor variables visual features **Video Characteristics + Scene Dynamics + Rigidity + Text + Color** and the number of views in 24 hours, for the milestone of 100,000 views. In terms of Kappa, the model obtained with this configuration achieved the best result than all the other

Table 2. Results and configurations of the 21 best models among the 105 models described in Section 4.3, selected based on the Kappa, Sensitivity, and Positive Predictive metrics. The model with the best result based on the Kappa metric is highlighted in bold.

Features	Views	Kappa	Sensitivity	Pos Pred Value	mtry
V	100k	0.7153	0.8716	0.8962	1
V	1m	0.5216	0.5416	0.6429	1
D	100k	0.4911	0.9144	0.7531	1
R	100k	0.4567	0.8471	0.7589	1
VC	100k	0.7154	0.8991	0.8776	8
VF	100k	0.7077	0.8685	0.8931	2
VF	1m	0.5216	0.5301	0.6567	1
VT	100k	0.7180	0.8807	0.8916	1
VT	750k	0.5427	0.5200	0.7429	2
VT	1m	0.5164	0.5181	0.6615	1
VD	100k	0.7243	0.8930	0.8875	1
VR	100k	0.7090	0.8869	0.8815	2
VR	750k	0.5691	0.5100	0.8226	1
VDF	100k	0.6892	0.8838	0.8705	2
VDT	100k	0.6897	0.8807	0.8727	2
VDR	100k	0.6805	0.8869	0.8631	3
VDRC	100k	0.4764	0.9021	0.8832	13
VDRT	100k	0.6901	0.8777	0.8750	7
VDRTC	**100k**	**0.7276**	**0.8991**	**0.8855**	**18**
VDRTF	100k	0.7194	0.8991	0.8165	5
VDRTFC	100k	0.7235	0.8991	0.8829	19

Table 3. Results of the 20 models generated with the of the number of views in the first hours after publication as predictor variables alone, described in Section 4.3. The model with the best result based on the Kappa metric is highlighted in bold.

Hours	Views	Kappa	Sensitivity	Pos Pred Value	mtry
1	10k	0.5474	0.9704	0.9665	1
1	100k	0.7004	0.9174	0.8571	1
1	500k	0.6202	0.7132	0.7077	1
1	750k	0.5795	0.6200	0.6889	1
1	1m	0.5459	0.5904	0.6364	1
6	10k	0.5744	0.9763	0.9667	1
6	100k	0.8702	0.9450	0.9508	1
6	500k	0.6794	0.7907	0.7286	1
6	750k	0.7241	0.7500	0.7979	1
6	1m	0.6479	0.7470	0.6667	1
12	10k	0.7407	0.9802	0.9822	1
12	100k	0.8550	0.9388	0.9174	1
12	500k	0.8396	0.8527	0.9016	1
12	750k	0.8245	0.8700	0.8447	1
12	1m	0.7467	0.7952	0.7765	1
24	10k	0.7680	0.9822	0.9842	1
24	**100k**	**0.9088**	**0.9511**	**0.9749**	**1**
24	500k	0.8485	0.8915	0.8779	1
24	750k	0.8272	0.8500	0.8673	1
24	1m	0.8249	0.8675	0.8372	1

models presented. Furthermore, it is possible to observe that models that use 1, 12 and 24 hours have better results combining the visual features than using only the number of views as predictors, comparing the results of Table 3 with Table 4.

5.4 Comparision

Using **only visual features**, Dalmoro and Musse [2021] obtained the best results in the model that used milestone 100,000 views and Video Characteristics and Rigidity as predictor variables, achieving a Kappa of 0.7324, sensitivity of 0.8930 and a positive predictive value of 0.8930. Comparing their results with the results of the present work, where we obteained Kappa of 0.7276, sensitivity of 0.8991 and positive predictive value of 0.8855 (Table 2), we can consider that the two techniques, SVM and Random Forest, had similar results, even though in Dalmoro and Musse [2021] the Kappa metric and the positive predictive value were slightly higher. In Trzciński and Rokita [2017], the best result presented using only visual features was the complete model,

Table 4. Results and configurations of models retrained with the visual features and milestones of the best model, according to Table 2, including the number of views in the first hours after the video was published. The best result based on the Kappa metric is highlighted in bold.

Feature	Hours	Views	Kappa	Sensitivity	Pos Pred Value	mtry
VDRTC	1	100k	0.8250	0.9205	0.9377	20
VDRTC	6	100k	0.8669	0.9358	0.9563	8
VDRTC	12	100k	0.9132	0.9388	0.9903	12
VDRTC	**24**	**100k**	**0.9358**	**0.9480**	**1.0000**	**20**

that is, the model that used all the visual features, which reached 0.23 in Spearman's correlation. Even though our technique and Trzciński and Rokita's method are different approaches (regression and classification), which makes it difficult to compare the results, we believe we have obtained better results in the model that uses only visual features. While they got 0.23 Spearman's correlation, we got Kappa of 0.7276, according to Table 2.

Using **only the number of views**, the results of our work and Trzciński and Rokita [2017] were better than using only visual features. This is because this is information directly correlated with the variable we are trying to predict. While our work the best model achieved a Kappa of 0.9088, Trzciński and Rokita [2017] obtained a Spearman's correlation of 0.9301.

Using **visual features and number of views**, both our results and the results of Trzciński and Rokita [2017] were better than using only visual features and using only number of views. Trzciński and Rokita [2017] reaching a Spearman's correlation of 0.9311, while our work the best model achieved a Kappa of 0.9358. As previously mentioned, the difference in approach makes it difficult to compare both results directly, but we can say that the improvement in both work results using the combination of the two types of information indicates that both the visual features and the number of views are good information to predict the popularity of a video and that, when used together, they tend to achieve better results.

6 Final considerations

In this work, we explore the importance of visual features along with the number of early views as features to predict the popularity of videos posted online. Using Random Forest, we predict which of the 1,820 videos posted to Facebook had more than a certain number of views seven days after they were published based on visual features extracted from the videos and the number of early views. By using the number of early views to predict their future popularity, it is possible, for example, for ads to be included as soon as possible, maximizing the audience reached. Combining visual features with the number of views makes popularity predictions perform even better and with the possibility to be executed before. Our predictive model performed better when using Video Characteristics, Scene Dynamics, Rigidity, Text and Color features, in conjunction with the number of views 24 hours after publication, obtaining a Kappa of 0.9358, a sensitivity of 0.9480, and a positive predictive value of 1 . As future work, we suggest testing features such as brightness and saturation, increasing the dataset size and comparing the results with videos of other types, such as ads and video lessons.

Declarations

Authors' Contributions

All authors contributed to the writing of this article, read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data can be made available upon request.

References

About Facebook (2018). Facebook watch: What we've built and what's ahead. Online; accessed 13-Sep-2021. https://about.fb.com/news/2018/12/facebook-watch-what-weve-built-whats-ahead/.

About Facebook (2020). The evolution of facebook watch. Online; accessed 07-Nov-2022. https://about.fb.com/news/2020/09/the-evolution-of-facebook-watch/.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32. DOI: 10.1023/A:1010933404324.

Dalmoro, B. M. and Musse, S. R. (2021). Predicting popularity of facebook videos through visual features using support vector machine classifier. In *Anais do XLVIII Seminário Integrado de Software e Hardware (SEMISH 2021)*. Sociedade Brasileira de Computação - SBC. DOI: 10.5753/semish.2021.15815.

Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling*, volume 26. Springer.

Kuhn, M. *et al*. (2008). Building predictive models in r using the caret package. *Journal of statistical software*, 28(5):1–26.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Ouyang, S., Li, C., and Li, X. (2016). A peek into the future: Predicting the popularity of online videos. *IEEE Access*, 4:3026–3033. DOI: 10.1109/access.2016.2580911.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rodriguez, J. D., Perez, A., and Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):569–575. DOI: 10.1109/TPAMI.2009.187.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Trzciński, T. and Rokita, P. (2017). Predicting popularity of online videos using support vector regression. *IEEE Transactions on Multimedia*, 19(11):2561–2570.

Vieira, S. M., Kaymak, U., and Sousa, J. M. C. (2010). Cohen's kappa coefficient as a performance measure for feature selection. In *International Conference on Fuzzy Systems*, pages 1–8. DOI: 10.1109/FUZZY.2010.5584447.

Youtube for Press (2022). Youtube by the numbers. Online; accessed 07-Nov-2022. https://blog.youtube/press/.

Zohourian, A., Sajedi, H., and Yavary, A. (2018). Popularity prediction of images and videos on instagram. In *2018 4th International Conference on Web Research (ICWR)*. IEEE. DOI: 10.1109/icwr.2018.8387246.