CBIE 2023
CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO
Uma escola para o futuro: Tecnologia e conectividade a serviço da educação

# Probabilistic classification of educational videos considering comments: an experimental analysis on Youtube

**Henrique C. F. B. Carvalho[1], Cristiano G. Pitangui[2], Fabiano A. Dorça[1], Catrine S. Oliveira[2], Eduardo A. C. Trindade[3], Alessandro V. Andrade[3], Luciana P. Assis[3]**

[1]Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brazil

[2]Departamento de Tecnologia e Eng. Civil, Computação e Humanidades
Universidade Federal de São João del-Rei (UFSJ) São João del-Rey – MG – Brazil

[3]Departamento de Computação
Universidade Federal dos Vales do Jequitinhonha e Mucuri – Diamantina – MG – Brazil

{henriquefbc, pitangui.cristiano, catrine.sntsoliveira}@gmail.com

fabianodor@ufu.br

{eduardo.trindade, alessandrovivas, lpassis}@ufvjm.edu.br

***Abstract.*** *Youtube is a constantly growing video platform that is massively used for teachers and students in the teaching and learning process. Some works point an important issue in Youtube search mechanism, as in many cases, the number of results returned by the platform is very large and not related to the search performed. In this sense, some works proposed methodologies to classy Youtube videos as educational or not to help in searching more specific educational content. This work develops a new methodology that probabilistically classify Youtube videos as educational or non-educational using its comments. Preliminary results show that comments can be used in order to probabilistically classify a video with high accuracy rates.*

## 1. Introduction

In Brazil, 9 out of 10 Youtube users access the platform with the intention of learning something, and more than half of them believe that it is the place where everything they want to see and learn can be found [Youtube 2019]. The Youtube platform, from an educational point of view, can be understood as a large repository of Learning Objects (LOs). LOs, in general, can be understood as any digital resource that can be reused in order to support learning, as long as it can be delivered over the network, such as images, videos, animations, texts and others [Wiley 2000]. Despite having an extensive collection of videos and providing content about diverse subjects, some problems related to the Youtube search engine can be identified.

As pointed in [Carvalho et al. 2020], one important issue in Youtube search mechanism is related to the results returned by the platform. In many cases, the number of results returned is very large, with many of them of low quality (considering educational

aspects) and/or not related to the search performed. In this sense, this considerable number of incorrect videos returned by the platform can be detrimental to teachers and students that use Youtube as a support for the teaching and learning process. In that regard, one natural try to surpass this problem is to classy Youtube videos as educational or non-educational, in order to support the platform in returning videos with educational content.

In [Carvalho et al. 2020] the authors analyzed 200 Youtube videos, being 100 educational and 100 non-educational. They identified relevant differences between the most frequent terms and words posted in the comments on educational and non-educational videos. The study showed that the comments posted by Youtube users have potential to be used in order to support categorization of videos. As a extension of the previous work, [Carvalho et al. 2022] analyses Youtube videos from an educational point of view and proposes a methodology for classifying them using its comments. In order to classify videos as educational or non-educational, the frequency of words used on comments of videos in this both categories was verified and the most frequent words are used for classification. The study demonstrates high accuracy when classifying an educational video and points out the main words used during its classification.

In this context, the present work, based on [Carvalho et al. 2020] and [Carvalho et al. 2022], develops a new methodology that probabilistically classify Youtube videos as educational or non-educational using its comments. Unlike in [Carvalho et al. 2022], the present work assigns a probability that a video belongs to one of the two classes under consideration. In this sense, this classification model is more flexible than the one adopted in [Carvalho et al. 2022], since it does not deterministically classify a video as educational or non-educational.

This probabilistic classification has a great advantage over the deterministic classification, in the sense that, in the real world, an Intelligent Tutoring System can recommend educational content that is classified as such with a certain probability. As an example, certain systems can be parameterized to recommend content classified with a minimum of 70% of being educational (less restrictive), or recommend content classified as 100% of being educational (highly restrictive). This same reasoning can be adopted in relation to the contents to be recommended. For example, some types of content may only be offered for videos that are rated as educational above a desired percentage threshold. On the other hand, other content may be recommended by videos rated with a less stringent percentage value. In addition, as another possibility, the same comments used to determine the probability of a video being educational or non-educational can also be used to measure the quality of the video through the use of sentiment analysis.

As an preliminary experiment, this work uses the dataset of videos and its comments from [Carvalho et al. 2020] in order to test the proposed classification methodology. Results showed that Random Forest and SVM classifiers are able do classify, with high accuracy rates, the video comments as educational or non-educational. In addition, it is showed that the classification of videos as educational or not, can be probabilistically performed through the use of the classification models from Random Forest and SVM algorithms. Thus, we point that this methodology has a great potential to classify Youtube videos in order to help students and teachers to find more appropriate learning contents on the platform.

The present work is organized as follows. The section 2 presents the main concepts considered in this work. The section 3 presents the main related works, and depicts how this work advances in the state-of-the-art. The section 4 describes the experimental methodology used in this work. The section 5 presents and discusses the obtained results. Finally, the section 6 presents final considerations, conclusions and future works.

## 2. Background

This section presents the main concepts related to this work.

### 2.1. Text Mining

Text Mining is the process that makes it possible to generate knowledge and extract relevant and non-trivial information from textual data. It is a multidisciplinary field that is based on Machine Learning, Data Mining, among others [Vijayarani et al. 2016, Jusoh and Alfawareh 2012]. It is similar to Data Mining techniques except that the tools used are designed to work on unstructured and semi-structured data such as: HTML files, emails, text documents, among others [Sukanya and Biruntha 2012]. Text Mining basically works in 3 stages: data pre-processing; application of Machine Learning/Data Mining techniques; and text analysis.

The pre-processing stage consists of treating the text before performing the analysis and application of the techniques. This step consists of standardizing the text by removing stop words (such as special characters and numbers) and clustering similar terms (i.e, converting characters to lowercase, correcting spelling errors, expanding and collapsing words) [Hickman et al. 2020]. The application of learning techniques consist of using algorithms to process data. Algorithms for clustering, classifying, visualizing, summarizing and extracting information can be used [Sukanya and Biruntha 2012]. The text analysis consists of a step to analyze and identify the relevant information that was generated after the previous step, thus obtaining the relevant information and generating knowledge about the processed text [Sukanya and Biruntha 2012].

### 2.2. Machine Learning

Machine Learning (ML) can be defined as the field of study concerned with how to provide the computer with the ability to learn without being explicitly programmed [Wiederhold and McCarthy 1992]. It is the branch of Artificial Intelligence that uses techniques and algorithms in order to recognize patterns or improve its performance through its experience [Mitchell et al. 1997, Russell and Norvig 2002].

One of the most common way to acquire knowledge by Machine Learning techniques is in Supervised Learning. In Supervised Learning, the data are sent along with labels and classes to which the data belong, that is, the algorithm already has previous information about the data provided. In this type of learning, the algorithms are provided with “training” and “test” data. In this way, it is necessary divide the data set into these two distinct parts.

After classifying the data, it is necessary to verify the true capacity of the classifier to recognize the classes presented. One of the most used and recommended methods to estimate the true prediction of supervised learning classifiers is through k-fold cross-validation method. This method basically consists of dividing the database into k parts,

using k-1 parts for the training stage and 1 part for the test stage, repeating this process k times, and modifying the sets of data, training and testing each time. In general, k = 10 is adopted, but other values for k can also be used [Berrar 2019, Mitchell et al. 1997].

One of the simplest and most successful ways to classify data is though Decision trees. A tree represents a function that takes as input a set of attributes and returns a "decision". Its decision is reached by executing a sequence of tests [Russell and Norvig 2002] . Each internal node in the tree corresponds to a test of the value of one of the input attributes. Each tree node is a test of some attribute and each child corresponds to a possible value of that attribute. Each leaf node represents a final variable value for a given input variable represented by the root node to the leaf node [Kesavaraj and Sukumaran 2013, Allahyari et al. 2017]. Random forest is a decision tree and supervised learning classifier model. It is an ensemble model, that is, it is an algorithm that builds several decision trees and the prediction evaluation is given by the set of these trees. After generating a large number of trees, they vote for the most popular class [Breiman 2001]. Another widely used classifier is the Support Vector Machine (SVM). It is a supervised learning classifier which is based on ideas originated in the statistical learning theory of [Vapnik and Vapnik 1998].

## 3. Related Works

This section presents and analyses some relevant related works. It is noteworthy that few works were identified regarding the categorization of educational videos from Youtube.

In [Abu-El-Haija et al. 2016] the authors address the classification of Youtube videos in order to develop a multiple video classification system. The database used contains approximately 8 million videos, encompassing a total of 1.9 billion video frames, and 500 thousand hours of categorized videos. The research was carried out in two stages, namely: 1) the video labels were obtained through Knowledge Graph entities; 2) the videos were processed frame to frame and categorized by a pre-trained Convolutional Neural Network in ImageNet. ImageNet is a visual database with several objects/entities already classified. Through the processing of more than 50 years of videos, generating 2 billion frames, and more than 8 million videos that can be quickly modeled on a single machine, the contribution of the work points towards helping the development of research on video understanding. Despite the various categorization classes, a specific category for educational videos was not found in this work.

In [Thelwall 2018] the authors analyze comments of Youtube videos related to dance styles. The database used contains 36,702 videos. The work aims to identify, through the comments posted on the platform's videos, the types of dance, gender relations (male and female), feelings expressed, and discussions regarding dance styles. For this purpose, the method called Comment Term Frequency Comparison (CTFC) is used in an attempt to identify subtopics/sub-themes of the discussions on a given topic in the Youtube comments, gender issues, feelings, and relationship between topics. The method successfully defined several prevailing attitudes in men and women. A specific category for educational videos was not found in this work.

In [Carvalho et al. 2020] the authors analyzed 200 Youtube videos, being 100 educational and 100 non-educational. They identified relevant differences between the most frequent terms and words posted in the comments on educational and non-educational

videos. Terms such as “best teacher” and “great class” are only present in the list of terms most often found in comments on educational videos. Similarly, it is demonstrated that the radicals “thank”, “lesso” and “teach” appear frequently in comments on educational videos. The study showed that the comments posted by Youtube users have potential to be used in order to support categorization of videos.

In [Carvalho et al. 2022], Youtube videos are analyzed from an educational point of view and a methodology for classifying them using its comments is also presented. In order to classify videos as educational or non-educational, the frequency of words used in this both categories was verified and the most frequent words are used for classification. Eight datasets were created and each one has a different number of most frequent words. The number of most frequent words used to create the datasets were 10, 20, 40, 60, 80, 100, 200 and 8; the latter was selected through attribute selection techniques. Each video in a dataset was represented by its *id*, the frequencies that the words appeared in its comments, and its class (educational or non-educational). The classification of the videos were performed by RIPPER, J48, and Random Forest algorithms. The classifiers receives the dataset and predicts a class for a video through the frequency of the words on its comments. The study demonstrates high accuracy rates when classifying an educational video and points out the main words used during its classification.

In this context, the present work, based on [Carvalho et al. 2020] and [Carvalho et al. 2022], develops a new methodology to probabilistically classify Youtube videos as educational or non-educational using its comments. Unlike in [Carvalho et al. 2022], the present work assigns a probability that a video belongs to one of the two classes under consideration. In this sense, this classification model is more flexible than the one adopted in [Carvalho et al. 2022], since it does not deterministically classify a video as educational or non-educational.

The main goal of this work is to develop a “mechanism” in order to assist Youtube search engine in returning educational content for students and teachers when using the platform to support their teaching and learning process. In this sense, the proposed work may be easily integrated in existing Virtual Learning Environments, providing reuse of educational content from Youtube platform. Therefore, this work advances the state-of-the-art in the field of use and reuse of educational content considering videos from Youtube, which is largely used for teachers and students in teaching and learning process.

## 4. Proposed Approach

The methodology adopted for the development of this work consisted in three steps described as follows.

1. **Dataset acquisition**. This stage considered the videos from [Carvalho et al. 2020], and the analysis of the educational videos was conducted based on the following definition: “a specific product, produced with a didactic-pedagogical intention and that considers its reception context, especially the school and the classroom, thus being intrinsically different from documentary videos, interviews, reports, etc´´ [Gomes 2008].
   In total, 200 videos were considered, being 100 educational and 100 non-educational. These videos provided a considerable number of comments (approximately 160,000). The obtained comments are already pre-processed and

ready to be used. This pre-processing procedure consisted in the following steps [Carvalho et al. 2020]: data normalization, accent removal, special characters removal, single characters and numbers removal, stopwords removal, and morphological normalization (stemming).

The dataset was modeled to present only the comments and its classes. The comment class is associated with the same class from the video in which the comment appeared. In this sense, all comments that were taken from educational videos are considered (classified) as educational, and all comments from non-educational videos are considered as non-educational.

2. **Classification of comments.** Random Forest [Breiman 2001] and SVM [Vapnik and Vapnik 1998] classifiers were used in this work to build classification models to classify comments, either as educational or not. In this sense, all the 160,000 comments from [Carvalho et al. 2020] were classified. The 10-fold cross-validation procedure [Berrar 2019, Mitchell et al. 1997] was used to construct the classification models and to generate the classification results. Both Random Forest and SVM were set with their default parameters as implemented in Python *scikit-learn* library.
3. **Probabilistic classification of videos.** The proposed approach uses equation (1) to calculate, $PEdu_{vi}$, which is the probability of a video *vi* being educational,

$$PEdu_{vi} = \sum edu_{vi} / (\sum edu_{vi} + \sum non\_edu_{vi}) \quad (1)$$

where:

- $\sum edu_{vi}$ is the sum of all comments classified as educational in *vi*;
- $\sum non\_edu_{vi}$ is the sum of all comments classified as non-educational in *vi*.

Similarly, one may calculate, using (2), $PNotEdu_{vi}$, which is the probability of a video *vi* being non-educational,

$$PNotEdu_{vi} = 1 - PEdu_{vi} \quad (2)$$

We developed a system to perform this probabilistic classification. The system was coded in Python and used the libraries *unidecode*, *regex*, *nltk*, *string* and the stopwords corpus in Portuguese. The Machine Learning module was implemented using the *scikit-learn* library with Random Forest and SVM classifiers.

## 5. Results and Discussions

As described in the last section, the dataset with the 160,000 pre-processed comments with its classes (educational or non-educational) was used to build Random Forest and SVM classification models. The Table 1 shows examples of pre-processed comments from both classes, educational (edu) and non-educational (non-edu), used for training the classifiers.

The Table 2 shows the accuracy results achieved by the classifiers using the 10-fold cross-validation method. It is observed that high accuracy results are achieved by both Random Forest and SVM when classifying comments as educational or non-educational. In this sense, it is pointed out that the small improvement in the accuracy

**Table 1. Examples of pre-processed comments from both classes, educational (edu) and non-educational (non-edu), used for training the classifiers.**

| Comments | Class |
|---|---|
| car alucinadokkk melhor professorkkkkkkkkkkkk | edu |
| aul excel | edu |
| melhor profes hist | edu |
| ach kkkk | edu |
| coloqu veloc parec profes ta chap kkkkkkkkkkkkkkkkkkk | edu |
| sucess carr sol bom heranc esper vc alcanc sucess aind mand cd ai kkkkkkkkkkkk | non-edu |
| faz temp ouv music heranc desd sai ms procur alg grup ano menos agor ach ire ouv | non-edu |
| temp procur dvd top demal | non-edu |
| guau busqu dvd grup herenc much tiemp per fin lo pud ver vay en verdad gen principi fin heranc dvd complet en la voz jaim juni | non-edu |
| hist nao | non-edu |

Source: Created by the authors.

results' for SVM, in relation to the Random Forest, is not statistically significant, therefore, both classifiers, considering accuracy results, are equivalent in this dataset.

Once the classifiers proved to be effective in classifying comments as educational and non-educational, the next step taken was to use them to support the probabilistic classification of videos, as explained in the previous section. In this sense, we selected two videos from [Carvalho et al. 2020], namely, "Herança Autossômica" and "Engenharia de Software - Aula 01 - Modelos de processo de software e atividades de software", to be probabilistically classified as educational or non-educational.

The Table 3 presents the comments on the video "Herança Autossômica". Also, it presents the pre-processed comments and it's classes (edu or non-edu) in according to the SVM' classification model. It can be noted that all the 6 pre-processed comments are classified as educational, thus, using equation (1), the system calculates $PEdu_{vi}$ = 6/(6+0) = 1. Thus, the video is classified as educational with probability equal to 1.

The Table 4 presents the comments on the video "Engenharia de Software - Aula 01 - Modelos de processo de software e atividades de software". Also, it presents the pre-processed comments and it's classes (edu or non-edu) in according to the SVM' classification model. It can be noted that from all the 5 pre-processed comments, 3 are classified as educational and 2 are classified as non-educational, thus, using equation (1), the system calculates $PEdu_{vi}$ = 3/(3+2) = 0.6. Thus, the video is classified as educational with probability equal to 0.6.

Here, it is important to highlight the flexibility of the interpretation of the probabilistic classification performed by the proposed methodology. In relation to the first video, 100% of its' comments are classified as educational and, therefore, it is concluded that this video has a high probability of being educational. Regarding the second video, only 60% of its comments are classified as educational and, therefore, the classification

**Table 2. Accuracy results' in classifying comments for SVM and Random Forest.**

| Fold | Random Forest | SVM |
|---|---|---|
| #1 | 83,97% | 84,56% |
| #2 | 83,78% | 84,53% |
| #3 | 83,82% | 84,63% |
| #4 | 83,70% | 84,56% |
| #5 | 83,99% | 84,65% |
| #6 | 83,84% | 84,55% |
| #7 | 84,02% | 84,67% |
| #8 | 83,85% | 84,62% |
| #9 | 83,83% | 84,66% |
| #10 | 83,89% | 84,66% |
| **Average** | **83,87%** | **84,61%** |

Source: Created by the authors.

**Table 3. SVM's classification of comments on video "Herança Autossômica"**

| Comment | Processed comment | Class |
|---|---|---|
| 7:47 Falha na realidade bem ali. | falh real bem ali | edu |
| Muito boa a aula ajudou bastante! | boa aul ajud bast | edu |
| Ótima aula | otim aul | edu |
| Suas aulas são muito boas!!! <3 | aul boa | edu |
| Muito bom, parabéns! Custei achar um vídeo que tratasse desse assunto de uma forma mais fácil e dinâmica de entender. Obrigada pela aula! | bom parab cust ach vide trat dess assunt facil dinam entend obrig aul | edu |
| Obrigada pela aula! | obrig aul | edu |

Source: Created by the authors.

of this video as educational should be interpreted with some parsimony.

In this sense, an Intelligent Tutoring System can consider a probabilistic threshold to recommend a particular video. As an example, the system can be parameterized to recommend videos that are considered educational with a probability above 80%. For videos classified as educational with probability lower than this limit, the system may try to consider other metadata from the video, such as title, number of likes and dislikes, description, etc., in order to verify whether the video fits into the educational category.

The proposed approach considers the frequency of educational comments on a video in order to probabilistically classify it as educational or non-educational. In this sense, this methodology easily allows the test and use of other techniques and algorithms with the same objective. Therefore, it is important to state that the proposed approach is very flexible, as it allows the classifier to be easily changed without interfering in the developed system's architecture, providing an easy way to produce new experiments and advances.

**Table 4. SVM's classification of comments on video "Engenharia de Software - Aula 01 - Modelos de processo de software e atividades de software".**

| Comment | Processed comment | Class |
|---|---|---|
| O dia é 6 e eu pretendo entrar em engenharia da computação. | dia pret engenh computaca | edu |
| Grato pela aula | grat aul | edu |
| Assistindo essas aulas percebo que ”antigamente” os professores da Univesp davam aulas mesmo, bem preparadas, com comentários relevantes e demonstrando profundo conhecimento. As aulas atuais (2021) são uma chatice, com o professor passando e lendo os slides. Lamentável. | assist aul perceb antig profes univesp dav aul bem prepar comentari relev demonstr profund conhec aul atual chat profes pass lend slid lamenta | edu |
| Experiência e prática: Uma boa forma de reduzir o tempo de produção de um software. Algo que depende do engenheiro. :) | experienc pra boa reduz temp produca softw alg depend engenh | non-edu |
| Amando cada vez mais a Engenharia de Software. Esse curso tem tudo o que eu quero seguir profissionalmente. #VemUFC | am cad vez engenh softw curs tud quer segu profiss vemufc | non-edu |

Source: Created by the authors.

## 6. Final Considerations and Future Work

As previously pointed out by [Carvalho et al. 2020], the category Education on Youtube is not assertive enough to classify videos with educative content as educational. In this sense, some works developed classification methodologies to classify Youtube videos either as educational or non-educational, in order to help students and teachers to search more specific educative content in the teaching and learning process.

The present work proposed a new methodology that probabilistically classify Youtube videos either as educational or non-educational. The proposed methodology is based on the use of classification models that classify the comments on videos as educational or not, and then use the frequency of the classes of the comments to perform the probabilistic classification of a video.

Results obtained over a dataset of 160,000 comments showed that SVM and Random Forest classifiers are able do classify, with high accuracy rates, the video comments either as educational or non-educational. In addition, it is showed that the classification of videos as educational or not, can be probabilistically performed through the use of the classification models from SVM and Random Forest algorithms. The use of the classification probability allows systems such as Intelligent Tutoring Systems, among others, to use probability boundaries in order to recommend Youtube videos.

As further works, we will advance in the probabilistic classification of an extensive Youtube video dataset in order to deeply investigate the true potential of the proposed classification methodology, and implement sentiment analyses of the video comments aiming to quantify its quality as a measure for a possible recommendation of the video by an Intelligent Tutoring System.

## References


Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., and Vijayanarasimhan, S. (2016). Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., and Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*.

Berrar, D. (2019). Cross-validation. In Ranganathan, S., Gribskov, M., Nakai, K., and Schönbach, C., editors, *Encyclopedia of Bioinformatics and Computational Biology*, pages 542–545. Academic Press, Oxford, UK.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Carvalho, H., Pitangui, C., Trindade, E., Assis, L., Andrade, A., and de Souza, D. (2020). Categorização de vídeos educacionais do youtube por meio de comentários. *RENOTE*, 18(2):621–629.

Carvalho, H. C. F. B., Dorça, F. A., Pitangui, C. G., de Assis, L. P., Andrade, A. V., and Trindade, E. A. C. (2022). Classificação automática de vídeos educacionais por meio de comentários apoiada por técnicas de aprendizado de máquina: uma análise experimental utilizando o youtube. *Revista Brasileira de Informática na Educação*, 30:419–448.

Gomes, L. (2008). Vídeos didáticos: uma proposta de critérios para análise. *Revista Brasileira de Estudos Pedagógicos*, 89(223).

Hickman, L., Thapa, S., Tay, L., Cao, M., and Srinivasan, P. (2020). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, page 1094428120971683.

Jusoh, S. and Alfawareh, H. M. (2012). Techniques, applications and challenging issue in text mining. *International Journal of Computer Science Issues (IJCSI)*, 9(6):431.

Kesavaraj, G. and Sukumaran, S. (2013). A study on classification techniques in data mining. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)*, pages 1–7. IEEE.

Mitchell, T. M. et al. (1997). *Machine learning*. McGraw-hill New York, New York.

Russell, S. and Norvig, P. (2002). Artificial intelligence: a modern approach.

Sukanya, M. and Biruntha, S. (2012). Techniques on text mining. In *2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pages 269–271. IEEE.

Thelwall, M. (2018). Social media analytics for youtube comments: Potential and limitations. *International Journal of Social Research Methodology*, 21(3):303–316.

Vapnik, V. N. and Vapnik, V., editors (1998). *Statistical learning theory*, volume 1. Wiley New York, New York.

Vijayarani, S., Janani, R., et al. (2016). Text mining: open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, 3(1):37–47.

Wiederhold, G. and McCarthy, J. (1992). Arthur samuel: Pioneer in machine learning. *IBM Journal of Research and Development*, 36(3):329–331.

Wiley, D. A. (2000). *Learning object design and sequencing theory*. PhD thesis, Brigham Young University.

Youtube (2019). Youtube insights. Acesso em: 17 de Abril de 2019.