Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

Aprimorando a experiência educacional no Youtube: uma abordagem de aprendizado de máquina para classificar e recomendar vídeos educacionais

Mejorando la experiencia educativa en Youtube: un enfoque de aprendizaje automático para clasificar y recomendar vídeos educativos

Henrique Carlos Fonte Boa Carvalho
Fabiano Azevedo Dorça
Cristiano Grijó Pitangui
Alessandro Vivas Andrade
Luciana Pereira de Assis
Eduardo Augusto Costa Trindade

Abstract

The fast development of technology has revolutionized social interaction and enabled easy access to a vast amount of information. However, it is increasingly challenging to find relevant educational materials within the large volume of available data. This challenge has led to a significant waste of time for teachers and students in searching for high-quality educational resources. In this sense, the present work focuses on classifying educational videos on YouTube using Machine Learning models. The study extends a previous work that analyzed...
Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

YouTube videos and proposed a methodology for classifying them using their comments. The current study expands the dataset used in the previous work and employs Machine Learning algorithms such as Random Forest and Neural Networks, along with hyperparameter tuning techniques like Grid Search. Experimental results showed that a Convolutional Neural Network was able to differentiate educational videos from non-educational ones with an accuracy rate of 95.71%. This study highlights the potential of Convolutional Neural Networks in classifying educational content on YouTube, contributing to advances in the field of Machine Learning for educational purposes.

Keywords: Classification. Education. Learning Objects. Neural Networks. Machine Learning. Youtube.

Resumo
O rápido desenvolvimento da tecnologia revolucionou a interação social e possibilitou o acesso fácil a uma vasta quantidade de informações. No entanto, é cada vez mais desafiador encontrar materiais educacionais relevantes dentro do grande volume de dados disponíveis. Esse desafio levou a uma significativa perda de tempo por parte de professores e alunos na busca por recursos educacionais de alta qualidade. Nesse sentido, o presente trabalho concentra-se em classificar vídeos educacionais no YouTube usando modelos de Aprendizado de Máquina. O estudo estende um trabalho anterior que analisou vídeos do YouTube e propôs uma metodologia para classificá-los usando seus comentários. O estudo atual expande o conjunto de dados usado no trabalho anterior e emprega algoritmos de Aprendizado de Máquina, como Random Forest e Redes Neurais, juntamente com técnicas de ajuste de hiperparâmetros como a Busca em Grade (Grid Search). Resultados experimentais mostraram que uma Rede Neural Convolucional foi capaz de diferenciar vídeos educacionais de não educacionais com uma taxa de acurácia de 95.71%. Este estudo destaca o potencial das Redes Neurais Convolucionais na classificação de conteúdo educacional no YouTube, contribuindo para os avanços no campo do Aprendizado de Máquina para fins educacionais.


Resumen
El rápido desarrollo de la tecnología ha revolucionado la interacción social y ha permitido el acceso fácil a una vasta cantidad de información. Sin embargo, es cada vez más desafiante
Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

Introduction

The advancement of technology allows society to be increasingly connected, with access to various information, social interaction, among others (Nascimento et al., 2017). The increase in data on the networks has provided a huge advancement for education, allowing new content to be created and shared every minute. However, this vast collection of content presents advantages and disadvantages, as pointed out by Miranda (2004): "In the area of education, for example, although many materials are being created and made available, access to them becomes a tiring and often failed process."

The failure in the search process occurs largely due to the enormous amount of documents presented to the user, which makes it difficult and confusing to select the most relevant ones (Braga & Menezes, 2014). The excess of data and available content hinders the teaching-learning process, causing teachers and students to spend most of their time searching for content, rather than studying or teaching. In this sense, the easy location and use of materials is of paramount importance in this process (Vieira & Nunes, 2012). It is important

encontrar materiales educativos relevantes dentro del gran volumen de datos disponibles. Este desafío ha llevado a una significativa pérdida de tiempo por parte de profesores y estudiantes en la búsqueda de recursos educativos de alta calidad. En este sentido, el presente trabajo se centra en clasificar videos educativos en YouTube utilizando modelos de Aprendizaje Automático. El estudio extiende un trabajo anterior que analizó videos de YouTube y propuso una metodología para clasificarlos usando sus comentarios. El estudio actual expande el conjunto de datos utilizado en el trabajo anterior y emplea algoritmos de Aprendizaje Automático como Random Forest y Redes Neuronales, junto con técnicas de ajuste de hiperparámetros como la Búsqueda en Cuadrícula (Grid Search). Los resultados experimentales mostraron que una Red Neuronal Convolucional fue capaz de diferenciar videos educativos de no educativos con una tasa de exactitud del 95,71%. Este estudio destaca el potencial de las Redes Neuronales Convolucionales en la clasificación de contenido educativo en YouTube, contribuyendo a los avances en el campo del Aprendizaje Automático para fines educativos.

to develop effective methods to classify and recommend educational content to facilitate access to the most relevant material and optimize the teaching-learning process.

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, 2019a). 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.

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 classify Youtube videos as educational or non-educational, in order to support the platform in returning videos with educational content (Carvalho et al., 2020).

Considering this scenario, this work extends previous work Carvalho et al. (2021), in which we analyzed 200 Youtube videos, being 100 educational and 100 non-educational. We 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 aim of the present work is to develop and test machine learning models capable of achieving higher accuracy considering previous work. To do so, we improved the dataset used in the previous work, increasing it to 500 features, and employed
Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

algorithms such as Random Forest and Neural Networks and applied hyperparameter tuning techniques such as Grid Search. In addition, we created a Convolutional Neural Network model that achieved an average accuracy of 95.71% and a maximum accuracy of 96.47%, surpassing the previous model's accuracy of 91.30%.

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.

Background

This section presents the main fundamental concepts related to this work.

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 & 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 & 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., 2022). 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 & 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 & Biruntha, 2012).

A way to acquire knowledge by Machine Learning techniques is through 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, 1997).

One of the simplest and most successful ways to classify data is through 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 (Russel & Norvig, 2010). 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 a (Kesavaraj & 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).

Neural networks, deep learning, and convolutional neural networks are widely used classification techniques based on mathematical models that simulate the activity of the human brain to process information. Neural networks are a collection of interconnected processing units that work together to process input information and generate outputs. Deep learning is a type of machine learning that uses neural networks with multiple layers to learn and recognize patterns in input data. Convolutional neural networks, on the other hand, are a type of neural network specifically designed to process data with a grid-like structure, such as images. They are highly effective for tasks related to image recognition and computer vision in general, as they can extract spatial features from images through convolutional layers (Lecun et al., 2015). These techniques are implemented, tested and their results are compared in this work.
Related Works

This section covers the main works related to this research. It should be noted that few studies were identified regarding the categorization of educational videos on YouTube, and no articles were found that address the recommendation of Learning Objects (LOs) through the use of comments. In this sense, this section presents studies that address the topics of LO recommendation using Wikipedia and YouTube, as well as studies that use Machine Learning techniques with a focus on YouTube. Furthermore, it is pointed out that the related works were collected through search platforms such as Google, Google Scholar, and the Capes Journals Portal.

In Menolli et al. (2011) authors aim to generate LO content through Wikipedia, using semantic technologies and the Learning Object Metadata (LOM) standard with Web 2.0. In their proposal, the content inserted into the platform is accessed, and text mining is performed to extract and classify the content according to the LOM standard. Using this standard makes it possible to find the attributes and metadata of the page, generating an XML-schema with the worked metadata. They conclude about the necessity of this approach, as it facilitates the use of content, since wiki tools do not consider how the content will be used.

In Abu-El-Haija et al. (2016) authors address the classification of videos on YouTube aiming to develop a multiple video classification system. The database used includes approximately 8 million videos, encompassing a total of 1.9 billion frames, and 500,000 hours of categorized videos. The research was conducted in two stages: 1) video labels were obtained through Knowledge Graph entities; 2) videos were processed frame by frame and categorized by a pre-trained Convolutional Neural Network on ImageNet. ImageNet is a visual database with various pre-classified objects/entities. By processing more than 50 years of videos, generating 2 billion frames, and over 8 million videos that can be quickly modeled on a single machine, the work aims to assist in the development of research on video comprehension. Despite the various categories of classification, no specific category was found for educational videos. The work cites the category ``Jobs \& Education" which includes universities, classrooms, lectures, etc. Therefore, a video that includes images of a university campus, for example, will be categorized in this category, although it may not necessarily be an educational video.

According to Júnior and Dorça (2018), authors present an approach for creating and recommending LO through the Wikipedia platform. The approach is defined by three steps:
1) ontology enrichment through metadata from wiki sections; 2) LO recommendation using Set Cover Problem combined with Genetic Algorithm techniques; 3) use of CRUD operations (Create, Read, Update, Delete). The study concludes that the adopted approach solves the LO recommendation problem, returning high-quality solutions.

In Pinheiro et al. (2018) authors present the Easy Youtube, a System of OA Recommendation based on Youtube. The system operates in six steps, as follows: 1) query enrichment - establishment of predefined themes, registered by specialists; 2) video extraction - video search, which can be done through search or predefined themes; 3) preprocessing - text treatment (in Portuguese), with punctuation and space removal, etc.; 4) classification - use of an algorithm to classify videos considered educational and of quality; 5) recommendation engine - the system receives the videos considered "good" and classifies them; 6) feedback collector - the user evaluates the recommendation provided by the system through ratings from one to seven stars. The work indicates its main contributions in the points: 1) the developed Recommendation System can be used as a solution for various application domains; 2) the system served as a proof of concept to improve recommendations through Youtube's features such as user ratings and native video language. However, the presented work does not detail important research issues. For example, for the classification of videos considered quality, it is stated that a training set of 100 videos was used, which were evaluated by specialists and students on the topic "Object Orientation/Inheritance". However, it is not explained how this analysis was carried out, and which characteristics of the videos were considered. Another point that causes confusion is the statement that, due to the short deadline for the research, the focus of the work was on "some characteristics for the experiment". Such characteristics were not described.

In Thelwall (2018), in turn, authors analyze comments on YouTube videos related to dance styles. The database used contains 36,702 videos. The work aims to identify, through comments posted on the platform's videos, dance types, gender relations (male and female), expressed feelings, and discussions related to dance styles. The Comment Term Frequency Comparison (CTFC) method is used to identify subtopics/subthemes of discussions about a specific topic in YouTube comments, gender issues, feelings, and topic relationships. The method successfully defines various predominant attitudes in men and women. The 10 terms associated with men were: shit, fuck, shuffle, man, fucking, crip, dude, bro, shuffling, hardstyle. On the other hand, the 10 terms associated with women were: she, amazing, her, beautiful, cute, omg, belly, ballet, really, workout. The sentiment analysis provided plausible
Revista de Gestão e Secretariado – GeSec, V. 15, N. 4, P. 01-20, 2024
São José dos Pinhais, Paraná, Brasil.

Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

Ideas for why the dances were appreciated. The 10 most used positive terms were: please, nice, wow, beautiful, loved, job (e.g. nice/great/good job), pretty, hope, perfect, keep (going/up the good work/it up). On the other hand, the 10 most used negative terms were: shit, fuck, killed, stupid, wtf, hate, idiot, dislike, die, dead. The authors claim that the results could serve as a starting point for further analysis on the topic and that the research would highlight gender differences, feelings, and subtopics between dances. They believe that the method used could be useful for discussing specific YouTube phenomena on a large scale, as well as for providing initial exploratory analysis of a certain problem that had not been previously researched in other contexts.

In Afonso and Duque (2019) sentiment analysis was performed on comments from YouTube videos using supervised machine learning techniques. The study collected 918 comments from a video that features analysis and critiques of the movie "Batman versus Superman: Dawn of Justice." The collected comments were classified as positive, negative, or neutral. Three experiments were performed: 1) three polarity classes: positive, negative, and neutral; 2) two classes: negative and non-negative; 3) only comments that featured the reference "Batman vs Superman movie" were used, and negative and non-negative classes were considered. The SMO classifier (Sequential Minimal Optimization algorithm for training a support vector classifier) and an 8-fold cross-validation methodology were used. The maximum accuracy obtained was 81%. The authors suggest that it might be possible to increase accuracy by collecting comments from other videos.

In Amanda and Negara (2020) authors applied Machine Learning techniques to classify YouTube videos as either "Kesenian" or "Sains", which in Indonesian means "Art" and "Science", respectively. The authors used the platform's search engine with the words "Kesenian" and "Sains" and obtained their experimental data in the form of video links, titles, and descriptions. Three classifiers were used: Random Forest, SVM, and Naïve Bayes. The Naïve Bayes classifier had the best evaluation, with an accuracy of 88%, while the Random Forest and SVM classifiers had an accuracy of 82%.

The primary objective of our research is to create a machine learning model that can accurately classify educational content on YouTube. In this study, we sought to improve upon the best models previously developed by incorporating neural networks, and we compared their performance with the previously best-performing algorithm, the Random Forest. To achieve this, we developed three different neural network models: a simple neural network (NN), a deep neural network (DNN), and a convolutional neural network (CNN). We
evaluated the performance of these models against the Random Forest algorithm, which had shown the best results in previous studies.

Our findings demonstrate a significant improvement over the previous state-of-the-art in educational content classification on YouTube. In particular, the CNN model we developed achieved higher accuracy and outperformed the Random Forest algorithm. This highlights the potential of CNNs to enhance the accuracy of educational content classification on YouTube and advance the field of machine learning for educational purposes.

**Proposed Approach**

The methodology adopted for the development of this work consisted in three steps illustrated in Figure 1 and described as follows.

**Figure 1**

*Proposed Approach*

1. **Dataset Acquisition**: This step considered the videos from (Carvalho et al., 2022). In total, 500 videos were considered, being 250 educational and 250 non-educational. These videos provided a considerable number of comments (approximately 738,500).
The obtained comments are already pre-processed and ready to be used. This pre-processing procedure consisted in the following steps Carvalho et al. (2022): data normalization, accent removal, special characters removal, single characters and numbers removal, stopwords removal, and morphological normalization (stemming).

Two datasets were obtained: the first dataset modeled in Carvalho et al. (2022) containing 200 words, which was the dataset where the best results were obtained. This dataset contains, for each selected video, Id, the words in its comments, and its class (educational or non-educational). Figure 2 shows an example of how the dataset was developed in (Carvalho et al., 2022).

Figure 2

Model of the dataset using 10 words

<table>
<thead>
<tr>
<th>Id</th>
<th>Video</th>
<th>aul</th>
<th>profes</th>
<th>obrig</th>
<th>vide</th>
<th>aprend</th>
<th>pra</th>
<th>ta</th>
<th>vc</th>
<th>faz</th>
<th>music</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>4g9JTQ2B6oo</td>
<td>21</td>
<td>28</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>uWa2WLOveuQ</td>
<td>79</td>
<td>52</td>
<td>29</td>
<td>17</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td>0</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>ZlB6MzmpKls</td>
<td>27</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>qa3FsuhBG8</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>L$Q0KMnakU4</td>
<td>33</td>
<td>38</td>
<td>17</td>
<td>23</td>
<td>10</td>
<td>18</td>
<td>4</td>
<td>6</td>
<td>16</td>
<td>2</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>7NKihiwokyk</td>
<td>9</td>
<td>31</td>
<td>66</td>
<td>214</td>
<td>17</td>
<td>241</td>
<td>46</td>
<td>119</td>
<td>322</td>
<td>0</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>DSBHxReMrd</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>22</td>
<td>6</td>
<td>52</td>
<td>4</td>
<td>13</td>
<td>59</td>
<td>0</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>UkhSLsDgj4M</td>
<td>56</td>
<td>254</td>
<td>135</td>
<td>343</td>
<td>53</td>
<td>811</td>
<td>92</td>
<td>344</td>
<td>1266</td>
<td>29</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>CFvy6zSsOEc</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>51</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>PHigOihSSY</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>0</td>
<td>44</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

Source: Obtained from Carvalho et al. (2022).

The second dataset was modeled in Carvalho et al. (2021) 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. In possession of the second dataset that contained the ids, comments and ratings, we generated a dataset containing 500 words, following the same methodology as (Carvalho et al., 2022).

2. Tuning and Modeling: Random Forest, Neural Network, Deep Learning and Convolutional Neural Network classifiers were used in this work to build classification models to classify comments, either as educational or not.

We used GridSearch to adjust the hyperparameters of the Random Forest. GridSearch is a technique used to find the best combination of hyperparameters in a machine
learning model. This technique involves evaluating all possible combinations of hyperparameters on a pre-defined grid and selecting the one that produces the best result according to a pre-determined metric, such as accuracy or precision (Géron, 2022). The grid was modeled according to Figure 3:

**Figure 3**

*Grid Search*

```
grid = {"n_estimators": [10, 100, 200, 500, 1000, 1200],
       "max_depth": [None, 5, 10, 20, 30, 50, 100],
       "max_features": [None, "sqrt", "log2"],
       "min_samples_split": [2, 4, 6, 8, 10],
       "min_samples_leaf": [1, 2, 4]}
```

The hyperparameter n_estimators specifies the number of trees in the forest. The grid above tests 6 different values for n_estimators: 10, 100, 200, 500, 1000, and 1200.

The hyperparameter max_depth specifies the maximum depth of each tree in the forest. The grid above tests 7 different values for max_depth: None (without a maximum depth limit) and 6 integer values (5, 10, 20, 30, 50, and 100) that limit the maximum depth of the tree.

The hyperparameter max_features specifies the maximum number of features considered in each node split in the tree. The grid above tests 3 different values for max_features: None (considers all features), "sqrt" (considers the square root of the total number of features), and "log2" (considers the base-2 logarithm of the total number of features).

The hyperparameter min_samples_split specifies the minimum number of samples required to split an internal node in the tree. The grid above tests 5 different values for min_samples_split: 2, 4, 6, 8, and 10.

The hyperparameter min_samples_leaf specifies the minimum number of samples required to be considered a leaf node in the tree. The grid above tests 3 different values for min_samples_leaf: 1, 2, and 4.

In summary, this grid specifies a total of 6 x 7 x 3 x 5 x 3 = 1890 different hyperparameter combinations for the Random Forest, and will test each of these combinations to find the best hyperparameter combination for the problem at hand.
To create the Neural Networks, Tensorflow was used, and it was necessary to define the network architecture, which includes choosing the number of layers, the number of neurons in each layer, the activation function, and other parameters.

After conducting experiments and comparative analyses, it was observed that the CNN showed superior results from the beginning. Although the DNN did not achieve better performance, there are possibilities for improvement. Therefore, based on these results, the neural networks were created, with emphasis on the CNN, aiming to maximize the accuracy and efficiency of the classification model. After defining the model to be created, each neural network underwent a training process of 100 epochs.

The architecture of the Neural Network used in the experiment is shown in Figure 4. The Deep Neural Network used in the experiment is shown in Figure 5. Finally, the Convolutional Neural Network is shown in Figure 6.

**Figure 4**

*Neural Network*

```
model = Sequential()
model.add(Dense(500, activation='relu',
                input_shape=(200,)))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

This is a basic neural network with two layers: a fully connected layer with 500 units and a ReLU activation function, and a single output layer with a sigmoid activation function. This model is used for binary classification problems, where the goal is to predict a binary output based on a set of input features. The model is compiled with the binary cross-entropy loss function and the Adam optimizer.
Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

Figure 5

Deep Neural Network

```python
model = Sequential()
model.add(Dense(500, activation='relu',
                input_shape=(200,)))
model.add(Dropout(0.1))
model.add(Dense(10, activation='PReLU'))
model.add(Dense(10, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy',
              optimizer='adam', metrics=['accuracy'])
```

This is a deeper neural network with 5 layers, including one dropout layers to prevent overfitting. The first layer is a fully connected layer with 500 units and a ReLU activation function, followed by a dropout layer with a rate of 0.1. The next 2 layers also have PReLU and relu activations. The final layer is a single output layer with a sigmoid activation function. The model is compiled with the binary cross-entropy loss function and the Adam optimizer.

Figure 6

Convolutional Neural Network

```python
model = Sequential()
model.add(Conv2D(filters=128, kernel_size=1,
                  activation='relu', input_shape=(200,1)))
model.add(Dropout(0.2))
model.add(Conv2D(filters=64, kernel_size=1,
                 activation='PReLU'))
model.add(Dropout(0.2))
model.add(Conv2D(filters=32, kernel_size=1,
                 activation='relu'))
model.add(Dropout(0.2))
model.add(MaxPooling1D(pool_size=1))
model.add(Flatten())
model.add(Dense(200, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(20, activation='PReLU'))
model.add(Dropout(0.2))
model.add(Dense(18, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy',
              optimizer='adam', metrics=['accuracy'])
```
This is a convolutional neural network designed for 1D data, with three convolutional layers followed by dropout and max-pooling layers. The first layer has 128 filters, a kernel size of 1, and a ReLU activation function. The second layer has 64 filters, a kernel size of 1, and a Parametric ReLU activation function. The third layer has 32 filters, a kernel size of 1, and a ReLU activation function. Each convolutional layer is followed by a dropout layer with a rate of 0.2. The max-pooling layer has a pool size of 1, which reduces the size of the feature maps. The flattened output of the max-pooling layer is passed to two fully connected layers, with 200 and 10 units respectively, and both with dropout layers. The final output layer has a sigmoid activation function. The model is compiled with the binary cross-entropy loss function and the Adam optimizer.

3. **Validation:** The 10-fold cross-validation procedure Berrar (2019) and Mitchell (1997) was used to construct the classification models and to generate the classification results. We ran 10 experiments, each experiment is the average that the classifier got when running the 10-fold cross-validation.

The experiments were conducted using Python programming language and the Scikit-learn library for Random Forest, along with the Tensorflow framework for implementing neural networks. The Random Forest algorithm was executed with both its default parameters and with hyperparameter tuning using the Grid Search method to identify the optimal parameters for the model.

The next section shows and discusses the results achieved in this research.

**Results and Discussion**

This section presents and discusses the results obtained during the experiments. We will begin the discussion with the results from the dataset containing 200 words and the Grid Search returned by Random Forest for this dataset, and then we will address the results from the dataset containing 500 words.

The best hyperparameters identified by Grid Search in the dataset containing 200 words are presented in the Figure 7 and the Table 1 present the accuracy for 200 words.
Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

Figure 7

Best hyperparameters for 200 words

```
'max_depth': 30,
'max_features': 'log2',
'min_samples_leaf': 2,
'min_samples_split': 4,
'n_estimators': 200
```

Table 1

Accuracy for 200 words

<table>
<thead>
<tr>
<th>Execution</th>
<th>RF</th>
<th>GS RF</th>
<th>NN</th>
<th>DNN</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>89.84%</td>
<td>90.05%</td>
<td>93.17%</td>
<td>93.81%</td>
<td>95.45%</td>
</tr>
<tr>
<td>2</td>
<td>89.42%</td>
<td>90.25%</td>
<td>92.96%</td>
<td>94.83%</td>
<td>95.45%</td>
</tr>
<tr>
<td>3</td>
<td>89.63%</td>
<td>90.68%</td>
<td>92.14%</td>
<td>95.65%</td>
<td>95.45%</td>
</tr>
<tr>
<td>4</td>
<td>89.01%</td>
<td>91.29%</td>
<td>92.95%</td>
<td>94.84%</td>
<td>95.02%</td>
</tr>
<tr>
<td>5</td>
<td>90.05%</td>
<td>90.88%</td>
<td>93.58%</td>
<td>94.01%</td>
<td>94.61%</td>
</tr>
<tr>
<td>6</td>
<td>89.43%</td>
<td>90.68%</td>
<td>93.79%</td>
<td>94.61%</td>
<td>94.81%</td>
</tr>
<tr>
<td>7</td>
<td>89.22%</td>
<td>90.26%</td>
<td>93.78%</td>
<td>94.41%</td>
<td>94.43%</td>
</tr>
<tr>
<td>8</td>
<td>88.80%</td>
<td>90.67%</td>
<td>93.18%</td>
<td>94.20%</td>
<td>94.63%</td>
</tr>
<tr>
<td>9</td>
<td>88.82%</td>
<td>90.46%</td>
<td>93.58%</td>
<td>94.61%</td>
<td>94.42%</td>
</tr>
<tr>
<td>10</td>
<td>88.40%</td>
<td>91.91%</td>
<td>93.58%</td>
<td>95.03%</td>
<td>95.03%</td>
</tr>
<tr>
<td>Mean</td>
<td>89.26%</td>
<td>90.71%</td>
<td>93.31%</td>
<td>94.60%</td>
<td>94.93%</td>
</tr>
</tbody>
</table>

RF = Random Forest; GS RF= Grid Search Random Forest; NN = Neural Network; DNN = Deep Neural Network; CNN = Convolutional Neural Network.

The Random Forest algorithm obtained slightly lower accuracy values compared to those obtained in (Carvalho et al., 2022). However, during the execution of the Grid Search and with the defined hyperparameters, we achieved better accuracy, indicating that the default parameters of Random Forest in Weka and Python are slightly different.

It is worth noting that the simple architecture of the developed Neural Network presented excellent results, with an average accuracy of 93.31%. Additionally, the Deep Neural Network showed better performance, improving the results in 1.29% when compared with a Simple Neural Network. Finally, it is noteworthy that the Convolutional Neural Network presented the best results among the experiments, with an average accuracy of 94.93% and a maximum accuracy of 95.45%.

The Grid Search technique was applied to the dataset comprising 500 words, resulting in the identification of the optimal hyperparameters shown in Figure 8.
Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos

Figure 8

Best hyperparameters for 500 words

Table 2

Accuracy for 500 words

<table>
<thead>
<tr>
<th>Execution</th>
<th>RF</th>
<th>GS RF</th>
<th>NN</th>
<th>DNN</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>89.42%</td>
<td>91.71%</td>
<td>94.41%</td>
<td>94.62%</td>
<td>95.66%</td>
</tr>
<tr>
<td>2</td>
<td>90.88%</td>
<td>91.51%</td>
<td>95.24%</td>
<td>95.03%</td>
<td>94.83%</td>
</tr>
<tr>
<td>3</td>
<td>91.08%</td>
<td>91.51%</td>
<td>94.19%</td>
<td>95.44%</td>
<td>96.26%</td>
</tr>
<tr>
<td>4</td>
<td>90.27%</td>
<td>92.33%</td>
<td>95.03%</td>
<td>94.82%</td>
<td>95.86%</td>
</tr>
<tr>
<td>5</td>
<td>90.26%</td>
<td>91.09%</td>
<td>95.03%</td>
<td>94.83%</td>
<td>95.66%</td>
</tr>
<tr>
<td>6</td>
<td>90.27%</td>
<td>91.51%</td>
<td>94.83%</td>
<td>95.23%</td>
<td>95.24%</td>
</tr>
<tr>
<td>7</td>
<td>90.67%</td>
<td>91.71%</td>
<td>95.44%</td>
<td>95.85%</td>
<td>95.24%</td>
</tr>
<tr>
<td>8</td>
<td>89.85%</td>
<td>92.34%</td>
<td>95.26%</td>
<td>95.24%</td>
<td>96.47%</td>
</tr>
<tr>
<td>9</td>
<td>90.68%</td>
<td>91.51%</td>
<td>94.62%</td>
<td>94.81%</td>
<td>95.86%</td>
</tr>
<tr>
<td>10</td>
<td>90.46%</td>
<td>91.09%</td>
<td>95.02%</td>
<td>95.62%</td>
<td>96.05%</td>
</tr>
<tr>
<td>Mean</td>
<td>90.38%</td>
<td>91.63%</td>
<td>94.91%</td>
<td>95.15%</td>
<td>95.71%</td>
</tr>
</tbody>
</table>

RF = Random Forest; GS RF = Grid Search Random Forest; NN = Neural Network; DNN = Deep Neural Network; CNN = Convolutional Neural Network.

It is observed that, in general, the algorithms presented better results with the larger dataset, containing 500 words. Additionally, it is noticed that the Convolutional Neural Network presented excellent results, and was the best algorithm in both datasets. The Convolutional Neural Network surprised us by obtaining an average accuracy of 95.71% and an experiment with an accuracy of 96.47%.

Final Considerations and Future Work

Identifying educational videos with high accuracy is crucial to ensure that users have access to reliable and relevant content. With the vast amount of videos available on YouTube, it is often difficult for users to find the best educational videos. In addition, the low quality or inaccuracy of some videos can lead to misinformation and lack of interest in learning.
This way, a machine learning model with high accuracy is essential for recommending learning objects from YouTube. The vast amount of educational content available on YouTube can make it challenging to find the most relevant and high-quality materials. By developing a machine learning model capable of accurately categorizing educational videos, it becomes possible to optimize the teaching and learning process by reducing the time spent searching for relevant content. This not only saves time but also ensures that teachers and students are exposed to the best educational content available, thus improving the quality of the learning experience.

This study endeavors to enhance the findings from previous research Carvalho et al. (2022) by expanding the dataset to include 500 features, performing a comparative evaluation of the Random Forest method, and applying Neural Networks to secure improved accuracy. The results demonstrated within indicate the feasibility of crafting a machine learning model boasting an accuracy surpassing 95%, an impressive feat particularly considering the complexity of the data involved.

The analysis shows that, in the context of the model proposed, Neural Networks and the Convolutional Neural Network outperform the Random Forest, which had been the most effective classifier in prior applications. The objective of this research is to devise a model of high accuracy that is suitable for use within a recommendation system, albeit at the cost of a minor reduction in model interpretability.

This research progresses along four distinct pathways: Firstly, to enhance and validate the LOIS (Learning Objects Intelligent Search), a Recommendation System designed to support both educators and students, through the application of a new machine learning model. Secondly, improve the LOIS to incorporating metadata and comments for more accurate categorization and recommendation of top-quality videos, including sentiment analysis to recommend those of the highest quality. Thirdly, to employ Machine Learning techniques for the identification of videos pertinent to the educational level and subject matter being instructed. Lastly, to develop and validate an educational vocabulary derived from the analysis of collected comments.

References

Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos


Improving the educational experience on Youtube: a machine learning approach to classifying and recommending educational videos


Received: 02.01.2024
Accepted: 03.11.2024