XXIV SemeAd
Seminários em Administração

novembro de 2021
ISSN 2177-3866

FINANCIAL EDUCATION IN BRAZIL ON YOUTUBE: A content analysis based on machine learning with topic models

DANIEL CASTRO MOTA
UNIVERSIDADE FEDERAL DA BAHIA (UFBA)

ANTONIO FRANCISCO DE ALMEIDA DA SILVA JUNIOR
UNIVERSIDADE FEDERAL DA BAHIA (UFBA)

FINANCIAL EDUCATION IN BRAZIL ON YOUTUBE:

A content analysis based on machine learning with topic models

1. INTRODUCTION

Financial education is a topic that is becoming increasingly important worldwide. Rising life expectancy and declining birth rates have led to changes in social security and pension systems around the world. The focus of financial decisions shifted from institutions to individuals, imposing on workers the responsibility to save, invest, and spend wisely throughout their life cycle (Lusardi & Mitchell, 2011).

The use of digital media has proved to be a way to broaden access to financial education, as technology is increasingly widespread in people's lives and can be used as an ally to spread knowledge in financial education. According to the Brazilian Institute of Geography and Statistics, 69.9% of the Brazilian population has internet access, of which 81.8% use it to watch videos (IBGE, 2017). Additionally, 81% of Brazilian students say they look for information about money on the internet (OECD, 2020).

The emergence of web 2.0 has led platforms such as blogs, wikis, social media, and video sharing sites to redefine the media consumption landscape. Also, improved broadband speeds and increased use of mobile devices in recent years have made online video consumption part of many people's daily routine. In this new era, viewers are no longer passive consumers and become active participants (Welbourne & Grant, 2016). YouTube is a significant example of the web 2.0 phenomenon. The video-sharing platform was founded in 2005 and has grown rapidly to become one of the leading sites on the Internet. Its technical and social characteristics support the formation of a participatory culture among community members in which they can develop, interact, and learn (Chau, 2010).

Due to the scale and diverse nature of its content, YouTube is attracting researchers to analyze its content from a social science perspective. With content distributed across several thematic categories, YouTube has such large scale and diverse data that it makes it a promising research object. When it comes to financial education, the situation is no different. In recent years, several YouTube channels have emerged to address this topic.

One of the factors that make the platform so attractive to researchers is the sheer availability of metadata associated with each video, such as the number of views, comments, likes, or dislikes, as well as audio transcripts. YouTube offers access to videos posted on its platform with automatically generated audio transcripts. This capture is possible by combining Google's *automatic speech recognition* (ASR) technology and YouTube's subtitles system. According to the official YouTube blog , the improvements implemented in the ASR algorithms allowed for a 50% increase in the accuracy of transcripts, approaching the error level of transcripts made by humans (YouTube, 2017).

The primary purpose of this article is to analyze the supply of financial education on YouTube videos (measured by the amount of content posted by YouTubers) and the demand for financial education content by YouTube audience (measured by video view count). For this, we reviewed a theoretical background of financial education content, wich was then related to the empirical results obtained from YouTube. By doing so, it allowed us to establish the relationship between supply and demand, identifying which financial education topics are most addressed and the level of interest of users for these themes.

The empirical work involves documentary research of the audio transcripts of 9,609 YouTube videos from 25 channels, totaling approximately 1,507 hours in length. To analyze that amount of data as objectively as possible, we used probabilistic topic modeling, a text-mining technique based on machine learning. Topic modeling algorithms are statistical methods that analyze the words of the original texts to find out the themes that compose them and how they are connected.

This paper is structured in five sections, including this introduction. The second section is devoted to the theoretical background, reviewing significant attempts to categorize financial education knowledge to date. In the third section, we present methodological procedures and data collection and outline some topic modeling concepts. Finally, we present the results and final considerations.

2. THEORETICAL BACKGROUND

The purpose of this section is to contextualize financial education. Firstly, we will present the definitions of the concept of financial education, according to the literature on the subject. Next, we will discuss how financial education content is structured.

2.1. Financial Education Concept

There is no standardized conceptual definition of financial education in the literature, but a wide range of meanings and related terms. The terms "financial education", "financial literacy" and "financial knowledge" are often used interchangeably, although some works make distinctions between them (Birochi & Pozzebon, 2016; Huston, 2010; Stolper & Walter, 2017).

One of the most used definitions is that of OECD, which started an intergovernmental project in 2003 to establish universal principles of financial education. For this institution, financial education can be defined as:

> [...] the process by which individuals and societies improve their understanding of financial concepts and products so that, with clear information, training and guidance, they acquire the values and skills necessary to become aware of opportunities and the risks involved in them, and then make informed choices, know where to look for help, take other actions that improve their well-being, thereby contributing consistently to the formation of responsible, future-oriented individuals and societies. (OECD, 2012, p. 7)

According to Huston (2010), literacy measures the ability to understand and use information. In this way, financial literacy can be understood as an indicator of an individual's ability to understand and use personal finance information. It is a component of human capital and has two dimensions: one concerning financial knowledge and another concerning the application of this knowledge. In turn, financial education is an *input* designed to increase this human capital.

Financial literacy aims to influence behaviors that increase financial well-being, which can be defined as the perception of being able to sustain the desired standard of living and financial freedom, both now and in the future (Brüggen et al., 2017). According to Huston (2010), other influences can also affect financial behavior and, consequently, financial well-being, such as behavioral bias, culture, or economic conditions.

The definition brought by Remund (2010), based on an extensive review of the concept of financial literacy in several government studies and programs between 2000 and 2010, is also worth noting:

> Financial literacy is a measure of the degree to which a person understands key financial concepts and has the ability and confidence to manage personal finances through appropriate short-term decision-making and sound long-term financial planning, while taking into account consideration of everyday events and changes in economic conditions (Remund, 2010, p. 284).

Remund's (2010) definition introduces an important element when it comes to personal finance, which is intertemporal choice. That is, the ability of individuals to ponder short- and long-term decisions.

Hilgert et al. (2003) explored the connection between knowledge and behavior in household financial management, focusing on four basic skills: cash flow management, credit management, savings, and investment. The authors conclude that there is a statistical correlation between financial knowledge and these practices, that is, increasing knowledge and experience seems to lead to an improvement in daily financial management practices.

Financial education is critical not only in wealth building issues such as retirement savings and equity investment but also in matters that concern the passive side of the household budget, i.e., debt behavior and debt management. Debt knowledge and skill levels play a crucial role in explaining how individuals organize their finances (Lusardi & Scheresberg, 2013).

2.3. Dimensions of Financial Education

As we try to understand which dimensions, or categories, the financial education content is structured into, two studies are especially relevant. The first one is a review of seventy-one academic studies by Huston (2010), in which the author identifies four content areas used in the financial education literature:

- Money basics: time value of money, purchasing power, personal financial accounting concepts.
- Indebtedness: use of future resources today by credit cards, personal loans, and financing.
- Investment: save present resources for future use through savings accounts, stocks, bonds, or mutual funds.
- Resource Protection: Insurance and other risk management techniques.

In another paper, Remund (2010) did an extensive review of government studies and programs between 2000 and 2010 to list what he defines as the most commonly found "operational categories":

- Budget,
- Savings,
- Loans, and
- Investments.

Other dimensions also appeared in the review, although they do not fit perfectly into these categories, such as insurance purchase, protection against overdue loans, and real estate financing (Remund, 2010).

In general, the main goal in categorizing financial education content is to formulate questionnaires that allow the measurement of financial knowledge. This theme has been the subject of several studies in the last decade. Typical questions in financial literacy tests seek to assess knowledge about financial products (stocks, bonds, funds, real estate financing, among others), knowledge about basic financial concepts (inflation, risk diversification, value of money over time), and mathematical and numerical skills (numeracy) (Stolper & Walter, 2017).

A pioneering work in this regard was that of Lusardi and Mitchell (2008, 2011, 2014), who elaborated simple questions to assess the level of financial education. The questions aimed to measure understanding of three fundamental concepts:

- Ability to perform simple interest rate calculations,
- Knowledge of inflation and the value of money over time, and
- Knowledge of risk diversification.

These questions became a reference for further studies and research and were replicated, adapted, and refined. A relevant example is the *S&P Global FinLit Survey* (Klapper et al., 2015), which measured the financial literacy of respondents in 140 countries. This research listed four concepts considered fundamental to financial decision making: (a) risk diversification; (b) inflation; (c) Basic calculations (numeracy); and (d) compound interest.

According to Atkinson and Messy (2012), the measurement of financial literacy should contemplate three aspects related to personal finance: knowledge, behavior, and attitudes. In the knowledge field, financially skilled people should be familiar with the concepts of division, time value of money, interest payment, interest calculation, compound interest, risk and return, and risk diversification. Noteworthy desirable behaviors for financial well-being are such as the practice of conscious consumption, personal budgeting, saving, and avoiding borrowing to meet expenses. The last element concerns attitudes and preferences, an attribute in which people can be divided between those who prioritize short term goals and those who prioritize long term goals. If people prefer to prioritize short-term consumption, they are unlikely to maintain an emergency reserve or make long-term financial plans (Atkinson & Messy, 2012).

3. DATA AND METHODS

This paper aims to analyze the supply and demand for financial education content on YouTube to understand which topics are addressed and the level of interest of users for these themes. We will do this by associating the proposed framework with the topic model generated from the audio transcripts and the number of views of each video from YouTube's financial education channels.

To choose which channels to analyze, we searched for videos with the keywords "Financial Education" and "Personal Finance". We limited our analysis to videos from the top 25 most-viewed channels, published between January 2016 and December 2020. This initial sample totaled 11,555 videos and over 1 billion views (see table 1).

Table 1. List of Financial Education YouTube Brazilian Channels.

Channel	**Views** (thousands)		**Subscribers** (thousands)		**Vídeos**	
Me poupe!	360,486	*34.0%*	6,210	*28.6%*	767	*6.6%*
O Primo Rico	211,055	*19.9%*	4,970	*22.9%*	493	*4.3%*
De bens com a vida	83,258	*7.8%*	80	*0.4%*	238	*2.1%*
EconoMirna	56,554	*5.3%*	1,230	*5.7%*	615	*5.3%*
Jovens de Negócios	48,261	*4.5%*	1,430	*6.6%*	183	*1.6%*
Júlia Mendonça	30,057	*2.8%*	562	*2.6%*	575	*5.0%*
Patricia Lages - Dicas de Economia	28,862	*2.7%*	640	*2.9%*	337	*2.9%*
Gustavo Cerbasi	28,506	*2.7%*	888	*4.1%*	477	*4.1%*
Cléber Miranda - Educação Financeira	27,830	*2.6%*	454	*2.1%*	606	*5.2%*
Dinheiro Com Você - Por William Ribeiro	25,311	*2.4%*	642	*3.0%*	392	*3.4%*
Clube do Valor	21,705	*2.0%*	737	*3.4%*	542	*4.7%*
Economista Sincero	16,627	*1.6%*	400	*1.8%*	253	*2.2%*
Maiara Xavier	14,981	*1.4%*	281	*1.3%*	528	*4.6%*
Bruno Perini - Você MAIS Rico	14,978	*1.4%*	590	*2.7%*	283	*2.4%*
Rafael Seabra	14,956	*1.4%*	312	*1.4%*	463	*4.0%*
Papo de Bolsa	13,677	*1.3%*	352	*1.6%*	376	*3.3%*
1Bilhão Educação Financeira	12,892	*1.2%*	334	*1.5%*	583	*5.0%*
André Bona	10,329	*1.0%*	237	*1.1%*	695	*6.0%*
Excelência no Bolso	9,338	*0.9%*	253	*1.2%*	562	*4.9%*
Dinheiro à vista	8,264	*0.8%*	203	*0.9%*	460	*4.0%*
Carlos Sampaio	5,623	*0.5%*	119	*0.5%*	455	*3.9%*
Finanças Femininas	4,971	*0.5%*	154	*0.7%*	414	*3.6%*
GuiaInvest	4,950	*0.5%*	213	*1.0%*	539	*4.7%*
Eu Quero Investir	4,371	*0.4%*	164	*0.8%*	575	*5.0%*
Nath Finanças	3,258	*0.3%*	248	*1.1%*	144	*1.2%*
Total	1,061,100	*100%*	21,703	*100%*	*11,555*	*100%*

Source: Elaborated based on YouTube data. Access in July 2021.

We used *YouTube Data Tools* software (Rieder, 2015) to access Youtube API and get the full list of videos from each channel, as well as their associated metadata: video identifier (*VideoID)*, publish date, title, duration, views, likes, dislikes and number of comments. We then used the open source program *youtube-dl* (2021) to download the audio transcripts from each video, in text format.

Descriptive statistics for the videos are presented in table 2. Each channel has an average of 462 videos, with videos lasting an average of 14.5 minutes. Views and length have a heterogeneous picture: the most popular video had been watched more than 20 million times, while the least watched had only 38 views (even though there was a six-month lag between the last publishing date and our data retrieval, which means the videos were available enough time to be seen and recommended by YouTube algorithm).

Table 2. Summary Statistics of Videos.

	Average	Median	Max	Min
Video views	91,838	14,082	20,315,090	38
Duration (minutes)	14.5	9.2	201.7	0.3
Channel views	44,006,252	15,804,018	360,485,958	3,258,328
Channel subscriptions	868,124	352,000	6,210,000	80,100
Channel video count	462	477	767	144

Source: research data.

As to length, the longest video had a duration of more than three hours (it was the recording of a live event). To avoid influence from such unusually long videos, we filtered length by using the IQR method for outlier detection, which resulted in a final sample of 9.609 videos, with maximum length of 27 minutes and nearly 940 million total views.

These data were then manipulated in *R* programming language, using the packages *readtext, readr, tidyverse,* and *quanteda*. As a final adjustment to the transcript data, we removed all stopwords or ‘empty words’, (i.e., words that do not add meaning and are very common, such as articles, prepositions, and linking verbs) and lemmatized the remaining words (i.e., transform different forms of the same word to the normalized form of the word).

3.1. Topic Modelling

Probabilistic topic modeling is a text-mining technique based on machine learning. Its algorithms use statistical methods to analyze the words in a set of texts and discover the themes (called “topics”) that compose them, assuming that the words carry strong semantic information and that similarly themed documents will use a group of similar words. There is no need for prior human intervention for document labeling, as topics will emerge from the probabilistic model. All documents in a corpus share the same set of topics, but each document exhibits these topics in different proportions.

The most known topic model algorithm is the Dirichlet latent allocation (LDA) proposed by Blei et al. (2003). For this work, we used the Structural Topic Model (STM), through the *stm* package in R, developed by Roberts et al. (2019). One of the advantages of STM over LDA is that it can incorporate metadata into the model as covariates.

Some parameters must be set before running the desired model. One of the biggest challenges when working with probabilistic topic modeling is selecting the number of topics to generate. Choosing a model with fewer topics will produce excessively generic results while choosing a model with many topics will result in an exaggerated division into many small and highly similar topics (Greene et al., 2014).

Roberts et al. (2014) suggest considering the criteria of semantic coherence and exclusivity in selecting the number of topics. Semantic coherence indicates how often words occur together for each topic. Exclusivity, as its name implies, measures how unique each word is to each topic in a set of topics. Ideally, we want to maximize both measures, but they go in opposite directions: more topics make each topic more specific, increasing semantic coherence. At the same time, the more topics, the less likely it is that words will be unique to a topic, thus decreasing exclusivity. The recommendation of Roberts et al. (2014) is to compare the exclusivity and semantic coherence of the candidate models and to select one model at the threshold where no model strictly dominates the other in terms of semantic coherence and exclusivity. Then we can select a model randomly or choose the one that is considered most appropriate in the researcher's judgment. After analyzing models with 20, 30, 35, 45, and 50 topics, we ended opting for the model with 35 topics, as it was the clearest in interpretation compared to the others.

Once the parameters were defined, the STM algorithm was applied to the document *corpus*, generating a probabilistic model that was then associated with the proposed categories in our framework. We used the probability associated with each topic as an indicator of content supply. For the content demand, we used video view counts, weighted by the probability associated with each topic for each video.

4. RESULTS AND DISCUSSION

The main outputs of a topic model are (a) the listing of words most likely to occur on each topic, and (b) the probability of occurrence of each topic on the corpus. Since it is common for the most frequent words to repeat across multiple topics, we also used the frequency-exclusivity (FREX) scoring method, as proposed by Roberts et al. (2014). This method calculates the harmonic mean frequency of a term under a topic with the exclusivity for that topic, ensuring that the chosen terms are, at the same time, frequent and exclusive, which helps to provide more semantically meaningful words.

Based on the listing of main words and justified by the theoretical background discussed in section 2, we assigned a name to each topic. For the most part, the topics presented good semantic coherence. Table 3 shows seven relevant keywords for each topic, along with their assigned names and probability of occurrence. Five topics were marked as "undefined" and eliminated from the analysis, as they aggregated words narrative expressions, terms, and jargon used by youtubers*,* and did not seem relevant to the purpose of this paper.

Table 3. Topic names, keywords and probability for each topic.

Topic/Theme	Relevant Keywords[1]	Probability
Investment (bonds and fixed income)	*selic, bond, treasury, savings, maturity, redeem, yield*	4.6%
Savings	*money, reserve, earn, emergency, save, mistake, salary*	4.4%
Calculations	*thousand, real, cent, month, million, calculation, compound*	4.3%
Undefined	*woman, nath, patricia, armored, blog, female, thanks*	4.2%
Undefined	*boy, nati, wedding, nat, natalia, party, love, mother, cry, husband*	3.9%
Investment (basics)	*broker, profile, invest, fixed, conservative, risk, aggressive*	3.8%
Entrepreneurship	*book, college, undertake, entrepreneurship, career, profession, develop*	3.8%
Financial behaviour (productivity, goals)	*goal, challenge, organize, accomplish, objective, routine, productive*	3.8%
Financial behaviour (motivational)	*prosperity, thoughts, brain, wealth, success, feeling, belief*	3.5%
News (economy)	*drop, crisis, rise, scenario, recover, market, cycle*	3.4%
Financial behaviour (family)	*children, dream, education, son, father, family, financial*	3.3%
Undefined	*andre, blog, board, nice, bona, perfect, decision*	3.3%
Financial behaviour (motivational)	*freedom, achieve, independence, millionaire, passive, mindset, purchase*	3.3%
Consumption tips (budgeting)	*save, spend, budget, supermarket, expenses, item,consumption*	3.3%
Investment (stock market)	*shareholder, dividend, company, growth, partner, distribute, profit*	3.0%
Banking products and consumer protection	*bank, inter, transfer, ted, digital, institution, fgc, pix, central, nubank*	2.8%
Undefined	*billionaire, fabricio, telegram, draft, shirt, punch, billion*	2.8%
Loans and debt management	*debt, check, financing, interest, pay off, score, loan*	2.7%
Retirement	*pension, retirement, private, plan, insurance, contribute, life, future*	2.7%
Consumption tips (clothing)	*cook, clothes, hair, wash, parts, make-up, pants*	2.7%
News (politics)	*elections, candidate, politic, population, president, bolsonaro, gdp*	2.6%
Investment (real estate funds)	*fund, real estate, shares, multimarket, manager, shareholder*	2.5%
Entrepreneurship	*entrepreneur, client, product, marketing, service, sales, business*	2.4%
Social benefits (emergency aid)	*aid, emergency, caixa, congress, register, pandemic, approve*	2.3%
Investment (strategies)	*club, ramiro, strategy, invest, allocation, return, method, portfolio*	2.2%
Investment (companies)	*luiza, magazine, retail, quarter, sector, preferred, ordinary*	2.1%
Taxes	*file, tax, underwriting, fill, right, invoice, income*	2.1%
Undefined	*anderson, excellence, pocket, goncalves, content, student, teacher*	2.1%
Investment (trading)	*trade, operate, day, market, analysis, chart, trend*	1.8%
Credit card	*card, miles, bill, annuity, friday, black, credit*	1.8%
Consumption tips (travel)	*car, travel, trip, ticket, driver, plane, hotel*	1.8%
Investment (international)	*etf, ibovespa, abroad, dollar, stocks, bdr, american*	1.8%
Social benefits (FGTS)	*fgts, anniversary, withdraw, release, balance, loan, dismiss*	1.8%
Investment (Currency exchange)	*bitcoin, currency, gold, pyramid, cryptocurrency, bubble, scam*	1.6%
Housing market	*house, rent, apartment, finance, condominium, land, building*	1.5%

Source: research data.

4.1. Measuring supply and demand

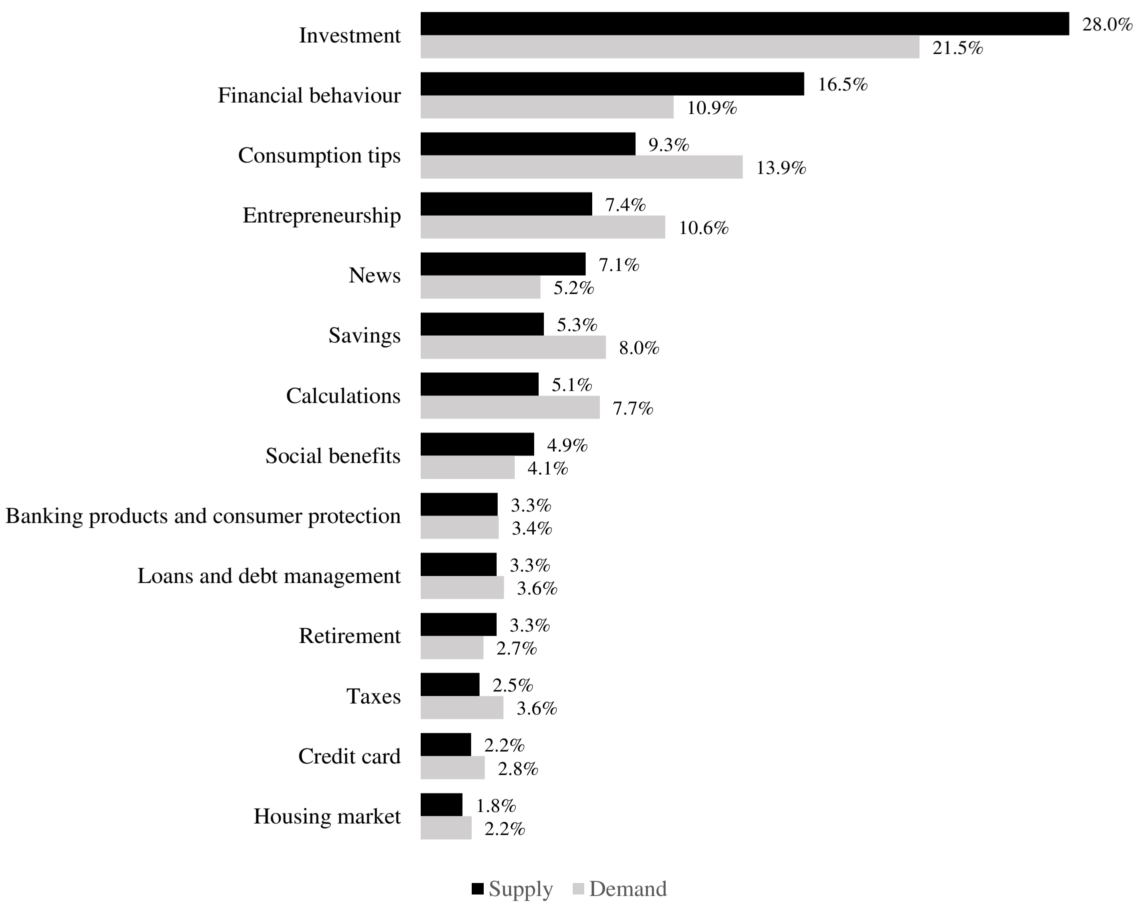
The probability of occurrence shown in table 3, also known as parameter θ, can be interpreted as how often each topic appears in the document corpus. We will use this parameter as a proxy for the financial education content supply on YouTube.

To measure content demand, we will use YouTube video view count, weighting the number of views of each video by the topic-document proportions (coefficient θ). The view for each topic is calculated as the sum of the product between the ratio of topic t in video i and the number of views in video i. By dividing the number of views per topic by the total view count, we obtain the proportion of the demand for each topic:

$$\frac{\sum_{i=1}^{n}(\theta_{t,i} \times views_i)}{\sum views}$$

As some topics had similar themes, we chose to aggregate them into categories for analysis. Figure 1 presents the comparison between supply and demand for each category.

Figure 1. Supply and demand of financial education content.



Source: research data.

The topics with the highest proportion of content produced by youtubers are those related to investment (28%). Videos with these topics talk about the stock market, performance of listed companies, real estate funds, bonds, and give general tips and strategies for operating in the financial market. Despite being a theme with a lot of content available, and the most demanded by users, it attracts proportionally less attention from users (21,5%). Topics that deal with financial behavior tips and news are also proportionately less viewed.

On the other hand, topics related to consumption (tips to save on shopping, take advantage of promotions and spend less on daily expenses), have demand higher than supply. This is also the case for topics related to entrepreneurship, savings, and financial calculations.

What seems to happen is that issues related to the short term, that is, more linked to people's day-to-day needs (“*How can I spend less?*”, “*How do I save money?*”, “*How do I open a business?*”), attract more views and are simpler to explain. Therefore, they have proportionally greater popularity. In the case of topics more focused on complex and long-term issues (“*How does this investment work?*”, “*Where can I invest my money?*”, “*How does the political scenario affect my investments?*”, “*How can I change my behavior towards money?*”), in addition to being less popular, require more video time to be explained to the audience.

5. FINAL CONSIDERATIONS

This article aimed to analyze the supply of financial education content on YouTube and its demand by users. For this purpose, we used an innovative approach based on topic modeling to analyze and label over 1,500 hours of audio transcriptions. We proposed a categorization of this content based on the theoretical background on financial education, and observed good adherence between the empirical result generated by this probabilistic model and the proposed categories.

The analysis of content offered and demanded allows us to infer that Brazilians looking for financial education content on YouTube are more concerned with less sophisticated and more short-term topics. YouTubers, on the other hand, offer many videos on investments, although the number of views is relatively lower. People watching personal finance videos on YouTube seem more concerned with balancing their budgets and improving their consumption patterns. These are commendable concerns, but they indicate a warning situation, given the increasing responsibility that individuals will have to take for their financial decisions throughout their life cycle. Nevertheless, this result is consistent with the degree of development of the Brazilian economy, which is a country with low saving rates, low average income, and high social inequality. Many Brazilians depend on social security and state assistance.

The contribution of this paper is twofold. First, it contributes to the literature in the field of financial education by consolidating a category structure for financial education content, validated empirically by topic modeling. Moreover, these results may contribute to the development of public policies that make the population aware of the importance of long-term financial planning. For instance, understanding public response to financial education themes may help to improve courseware and other teaching material.

Finally, it is worth highlighting the limitations of this research. Since it was based on audio transcripts, transcription errors can affect the accuracy of the model. However, we observed a high level of semantic coherence in the generated topics. Although much of the Brazilian population has access to the Internet, it is imperative to highlight that the results cannot be generalized to the Brazilian population, since the observed sample consisted solely of Brazilian users of the YouTube platform.

This work may also lead to future research. One possibility is to conduct a more in-depth qualitative study of the content of the videos, aiming to evaluate the adequacy of what YouTubers say concerning each of the topics addressed. The experience of these financial education content producers is quite diverse; some are certified professionals, while others are amateur investors sharing their experiences, so the quality and degree of exemption can vary greatly. Along these lines, it would also be interesting to study conflicts of interest on the content produced, since some YouTubers may be sponsored by financial institutions.

6. REFERENCES

Atkinson, A., & Messy, F. (2012). Measuring Financial Literacy: Results of the OECD / International Network on Financial Education (INFE) Pilot Study. *OECD Working Papers on Finance, Insurance and Private Pensions*, *15*, 1–73. https://doi.org/10.1787/5k9csfs90fr4-en

Birochi, R., & Pozzebon, M. (2016). Improving financial inclusion: Towards a critical financial education framework. *Revista de Administração de Empresas*, *56*(3), 266–287. https://doi.org/10.1590/S0034-759020160302

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Brüggen, E. C., Hogreve, J., Holmlund, M., Kabadayi, S., & Löfgren, M. (2017). Financial well-being: A conceptualization and research agenda. *Journal of Business Research*, *79*, 228–237. https://doi.org/10.1016/j.jbusres.2017.03.013

Chau, C. (2010). YouTube as a participatory culture. *New Directions for Youth Development*, *2010*(128), 65–74. https://doi.org/10.1002/yd.376

Greene, D., O'Callaghan, D., & Cunningham, P. (2014). How Many Topics? Stability Analysis for Topic Models. *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, *v.8724*, 498–513. https://doi.org/10.1007/978-3-662-44848-9_32

Hilgert, M. A., Hogarth, J. M., & Beverly, S. G. (2003). Household financial management: The connection between knowledge and behavior. *Federal Reserve Bulletin*, *Jul*, 309–322.

Huston, S. J. (2010). Measuring Financial Literacy. *Journal of Consumer Affairs*, *44*(2), 296–316. https://doi.org/10.1111/j.1745-6606.2010.01170.x

IBGE. (2017). *PNAD Contínua: Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal* (p. 12 p.). IBGE. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101631

Klapper, L., Lusardi, A., & van Oudheusden, P. (2015). *Financial Literacy Around the World: Insights from the Standard & Poor's rating services global financial literacy survey* (p. 28). http://gflec.org/wp-content/uploads/2015/11/Finlit_paper_16_F2_singles.pdf

Lusardi, A., & Mitchell, O. S. (2008). Planning and Financial Literacy: How Do Women Fare? *American Economic Review*, *98*(2), 413–417. https://doi.org/10.1257/aer.98.2.413

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy around the world: An overview. *Journal of Pension Economics & Finance*, *10*(4), 497–508. https://doi.org/10.1017/S1474747211000448

Lusardi, A., & Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, *52*(1), 5–44. https://doi.org/10.1257/jel.52.1.5

Lusardi, A., & Scheresberg, C. de B. (2013). *Financial Literacy and High-Cost Borrowing in the United States* (Working Paper No. 18969). National Bureau of Economic Research. https://doi.org/10.3386/w18969

OECD (2012). *OECD/INFE High-level Principles on National Strategies for Financial Education—Financial Stability Board* (p. 20). OECD Publishing. http://www.oecd.org/daf/fin/financial-education/OECD_INFE_High_Level_Principles_National_Strategies_Financial_Education_APEC.pdf

OECD (2020). *PISA 2018 Results (Volume IV): Are Students Smart about Money?* OECD Publishing. Paris. https://doi.org/10.1787/48ebd1ba-en

Remund, D. L. (2010). Financial Literacy Explicated: The Case for a Clearer Definition in an Increasingly Complex Economy. *Journal of Consumer Affairs*, *44*(2), 276–295. https://doi.org/10.1111/j.1745-6606.2010.01169.x

Rieder, B. (2015). *YouTube Data Tools* (1.12) [Computer software]. https://tools.digitalmethods.net/netvizz/youtube/

Roberts, M., Stewart, B., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. Journal of Statistical Software, 91(2), 1 - 40. doi:http://dx.doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Stolper, O., & Walter, A. (2017). Financial literacy, financial advice, and financial behavior. *Journal of Business Economics*, *87*(5), 581–643. https://doi.org/10.1007/s11573-017-0853-9

Welbourne, D. J., & Grant, W. J. (2016). Science communication on YouTube: Factors that affect channel and video popularity. *Public Understanding of Science*, *25*(6), 706–718. https://doi.org/10.1177/0963662515572068

YouTube. (2017, February 16). One billion captioned videos. *Official YouTube Blog*. https://youtube.googleblog.com/2017/02/one-billion-captioned-videos.html

Ytdl-org/youtube-dl. (2021). [Python]. youtube-dl. https://github.com/ytdl-org/youtube-dl (Original work published 2010)