



JUST ONE POST? FORECASTS OF DAILY SALES OF BEAUTY AND COSMETICS RETAIL COMPANIES BASED ON THE INFLUENCE OF SOCIAL MEDIA

 **Gabriel Rodrigo Gomes Pessanha**

PhD in Business
Federal University of Alfenas (UNIFAL) - Institute of Applied Social Sciences (ICSA)
Varginha, MG – Brazil.
gabriel.pessanha@unifal-mg.edu.br

 **Eduardo Almeida Soares**

Master
Lancaster University
United Kingdom
e.almeidasoares@lancaster.ac.uk

Objective: To study the relevance of *Instagram* posts in the construction of forecasting models for the variation of daily sales revenues for retail companies in the beauty and cosmetics sector.

Methodology: Time series of daily sales between the years 2017 and 2019 of 10 retail companies in the beauty and cosmetics sector were considered. Methods based on machine learning were used and the forecasting models were increased with numerical variables from the official profile of the company, from the posting made by the contracted digital influencer and the characteristics of the images posted by the digital influencer were included in the models.

Relevance and Originality: The study is innovative, as it goes beyond qualitative reflections on the theme and provides empirical evidence regarding the impacts on forecast accuracy from the inclusion of social media variables. A data fusion strategy (numerics and images) was also presented to forecast daily sales of retail companies in the beauty and cosmetics sector.

Main results: The models proved to be efficient in forecasting and the importance of the likes and engagement variables reinforces the idea that the identification and social reference generated by the ID are important aspects in the purchase decision process. It was found that the images are responsible for adding exclusive attributes that help in forecasting and understanding the patterns of the sales series.

Theoretical and methodological contributions: The study showed in a promising way the efficiency of methods based on machine learning in forecasting sales from *Instagram* data, especially with regard to the incorporation and extraction of image data.

Keywords: Social media. Images. Artificial intelligence. Sales forecasting. Digital marketing. Digital influencer.

How to cite the article

American Psychological Association (APA)

Pessanha, G. R. G., & Soares, E. A. (2021, Oct. /Dec.). Just one post? forecasts of daily sales of beauty and cosmetics retail companies based on the influence of social media. *Brazilian Journal of Marketing*, 20(4), 241-266.
<https://doi.org/10.5585/remark.v20i4.17914>.



Introduction

In the current competitive environment, planning becomes an important task to generate competitive advantages in the decision process. In this context, demand forecasting is essential for organizational planning, as it makes it allows us to understand the influences of internal and external factors on demand behavior and its future measurement (Chen & Lu, 2017; Kremer, Siemsen, & Thomas, 2016; Trapero, Kourentzes, & Fildes, 2015). Thus, it becomes an essential step for the development of strategies related to inventories, labor, purchasing, production plans, in addition to managing customer satisfaction and financial savings throughout the supply chain (Kremer et al., 2016; Fildes, Goodwin, & Önkal, 2019).

Numerous variables can influence demand dynamics, and therefore statistical models that only consist of historical sales data are often insufficient to produce accurate forecasts (Abolghasemi, Eshragh, Hurley, & Fahimnia, 2019). Uncertainty makes planning and control more difficult (Slack, Chambers, & Johnston, 2009; Hyndman & Athanasopoulos, 2014). According to Lawrence, Goodwin, O'Connor and Önkal (2006), although historical data are important for generating forecasts, they are often insufficient to generate accurate predictors, as models do not have the ability to update contextual aspects and incorporate new information.

Events related to the promotion of products and services are important and can increase the accuracy of forecasting models. Trapero et al. (2015) found that marketing campaigns significantly influence market demand. Several studies have shown that the incorporation of promotional marketing campaigns increases the accuracy of sales forecasts (Fildes & Goodwin, 2007; Goodwin, 2002). Although several studies have analyzed the impact of marketing campaigns on sales (Alalwan, 2018; Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016; Lamond et al., 2010; Ramanathan & Muiyldermans, 2011; Oliva & Watson, 2009; Huang, Fildes, & Soopramanien, 2014; Kourentzes & Petropoulos, 2016) Abolghasemi, Eshragh, Hurley and Fahimnia (2019) state that the task of quantifying the impact of promotions is still problematic for the practice of academic analysts and researchers, especially when it' has to do with social media marketing campaigns.

Social media has become an indispensable source of information and content (Lou & Yuan, 2019; Shiau, Dwivedi, & Lai, 2018). Furthermore, such platforms have proven effective in promoting marketing campaigns (Alalwan, 2018; De Vries & Carlson, 2014; Gensler, Völckner, Liu-Thompkins, & Wiertz, 2013). According to Sokolova and Kefi (2020), in addition to publications and fan pages, users actively generate and publish multimedia content,

including their opinions about brands and products. In this new process, digital influencers create compelling content to market products or services, targeting a specific audience niche.

Thus, the use of digital influencers for advertising and marketing campaigns is proving increasingly popular among companies, as, according to Childers, Lemon and Hoy (2019), digital influencers connect the public to brands, maintaining a channel of open communication with followers. Furthermore, some studies (Welbourne & Grant, 2016; Gillon, Aral, Lin, Mithas, & Zozulia, 2014; Lipizzi, Iandoli, & Marquez, 2015) demonstrate that advertising by digital influencers is more efficient than traditional professional advertising.

Other research has highlighted the value of social media in improving sales variation forecasting models (Liu, 2006; Lassen, Madsen, & Vatrapu, 2014; Kulkarni, Kannan, & Moe, 2012; Asur & Huberman, 2010; Gruhl, Guha, Kumar, Novak, & Tomkins, 2005). However, despite the effort to quantitatively understand the influence of digital platforms and their users on product demand, there is still no consensus on the impact of social networks and digital influencers in increasing sales and profitability. Thus, the general hypothesis of this work is based on the predictive power of posts made by digital influencers. In other words, it is believed that the characteristics of posts (Likes, comments, engagement and images) can be configured as predictive variables with the power to increase the accuracy of traditional models, based solely on historical sales behavior.

The purpose of this research is to study the relevance of Instagram posts in the construction of daily sales revenue variation forecast models for retail companies in the beauty and cosmetics industry.

In addition to the introductory section, the article is structured as follows: in section 2, the theoretical context in which the research was developed is presented; in the next section, the methodology used in the analyses and the procedures for comparing the prediction performance of the models used are described; in section 4, the analysis and discussion of the results are shown; final considerations and future work directions are presented in section 5; and, finally, section 6 brings the bibliographical references used throughout this work.

2 Theoretical reference

2.1 Social media and marketing – previous studies

In recent years, some studies have investigated the relationship between social media and consumer behavior (Alalwan, Rana, Dwivedi, & Algharabat, 2017; Hajli & Sims, 2015; Hajli, 2014). In order to raise the most cited studies of the last 5 years, a search was carried out in the SCOPUS database with the keywords “social media”, “digital influencers” and “sales”. Table 1 presents a summary of some previous work related to social media. The relationship between digital influencers (IDs) and purchasing decisions has been the subject of several academic research.

Jin (2018) analyzed the effects of Facebook on advertising through ANOVA, mediation analysis and Bootstrapping. The author conducted two experiments that confirm the justification theory in the endorsement of Facebook-based celebrity brands. The survey results indicate that third-party endorsement influences source credibility and ad effectiveness more positively than self-endorsement. In addition, social identification and engagement are significant metrics and moderators.

More and Lingam (2019) provided an overview of influencer identification and influence maximization. Using an epidemic infection model, the authors concluded that by identifying influential users on social media, different business strategies can be devised, and that influence maximization and social influence mining together will form the significant components to enable broad marketing viral through online social networks.

Audrezet, De Kerviler and Moulard (2018), through non-participatory observations and semi-structured interviews, studied the strategies that digital influencers (DIs) use to maintain their personal authenticity when partnering with brands. From the results, two authenticity management strategies emerged from the analysis: passionate and transparent authenticity. Finally, the authors proposed a four-way framework that provides the first conceptualization of how DIs can manage authenticity to deal with the tensions created by brand intrusion into their content.

Table 1

Some Previous Studies Related To The Theme

Author	Objective	Main results
Babić Rosário, Sotgiu, De Valck and Bijmolt (2016)	To conduct a meta-analysis to measure the impact of eWOM on sales of different products.	On average, eWOM is positively correlated with sales but its impact varies by product type, digital and metric platform used.
Kumar, Bezawada, Rishika, Janakiraman and Kannan (2016)	To study the effect of brand-generated content on social media on three key consumer metrics (spending, cross-purchasing, and consumer profitability).	Combined with television advertising and email marketing, the FGC had a positive effect on consumer behavior.
Lu, Fan and Zhou (2016)	To theorize the nature of the social aspect in the online social commerce (SC) market, proposing three variables of social presence.	Social presence factors based on social technologies have a significant positive effect on online commerce.
Erkan and Evans (2016)	To examine the effect of eWOM on social media on consumers' purchase intent.	The quality, credibility, usefulness and adoption of information, need for information and attitude towards the brand are key factors of eWOM in social media that influence the purchase intention of consumers.
De Veirman, Cauberghe and Hudders (2017)	To assess, based on the number of followers on Instagram, how digital influencers are seen and how they can change consumers' perception of a brand and the originality of a product.	Digital influencers (DIs) with a high number of followers tend to be better seen, but this does not always translate into greater opinion-forming power among their followers. Following too few profiles, however, can damage the image of a DI. Furthermore, very popular DIs tend to damage consumers' perception of the originality of a product or a brand.
Godey et al. (2016)	To study the effect of social media marketing on brand value creation and consumer behavior with respect to pioneering brands in the luxury goods industry.	The study demonstrates the relationships between marketing actions via social media and brand preference, premium prices and loyalty. Another contribution of the authors is the discovery that SMMEs have a positive impact on brand equity and their two main dimensions are: brand awareness and brand image.
Schivinski and Dabrowski (2016)	To analyze the effects of communication via social media on consumers' perception of brands.	Communication on social media via standard user positively influences both brand equity and attitude towards it, whereas communication via brand only affects attitude towards the brand. It has also been shown that brand equity and brand attitude have a positive influence on purchase intentions.

Author	Objective	Main results
Hudson, Huang, Roth and Madden (2016)	To evaluate how the use of social media influences consumer-brand relationships.	The use of social media is positively related to the quality of interaction between consumer and brand. This effect increases with brand personification.
Agnihotri, Dingus, Hu and Krush (2016)	To test the effects of the sellers' informational communication behaviors on social media usage and consumer satisfaction.	It was observed that the use of social media by sellers has an impact on informational communication behaviors, intensifying seller responsiveness and consumer satisfaction.
Lee, Hosanagar and Nair (2018)	Describe the effect of social media advertising on consumer engagement through Facebook data.	The inclusion of content related to a brand's personality, such as humor and emotion, is associated with higher levels of consumer engagement. Content with more direct information, such as prices and offers, is associated with lower levels of consumer engagement. On the other hand, a combination of the two types of content leads to higher levels of engagement.
Lou and Yuan (2019)	To understand the mechanism by which ID marketing affects consumers.	The informative value of content created by a DI, as well as its credibility, attractiveness and similarity to its followers are associated with a higher level of trust by followers in the posts advertising a DI.
Alalwan (2018)	To identify and test the main factors related to advertising via social media that have the power to predict purchase intentions.	The results indicated that the factor performance expectation, hedonic motivation, interactivity, informativeness and relevance perception had a significant effect on the prediction of purchase intentions.
Sokolova and Kefi (2020)	To investigate how the parasocial interaction (PSI) between audience and DI in the fashion and beauty areas of YouTube and Instagram, together with their perception of credibility, influence purchase intentions and how they relate to social and physical attractiveness and the homophily factor.	The homophily factor is positively associated with PSI, while physical attractiveness showed an opposite behavior, being negatively associated with PSI or showing no evidence of association. Both DI's and PSI's credibility were positively related to purchase intentions.
Arora, Bansal, Kandpal, Aswani and Dwivedi (2019)	To propose a mechanism to measure the index of a DI in the most popular social media, including Facebook, Twitter and Instagram.	Engagement, reach, sentiment and growth play a key role in determining a DI index.

Author	Objective	Main results
Schouten, Janssen and Verspaget (2020)	To investigate the difference between the impact of a brand/product recommendation made by a traditional celebrity and a social media DI.	Overall, the results showed that there is greater advertising effectiveness when a product/brand is endorsed by a DI compared to a traditional celebrity.
Araujo, Neijens and Vliegenthart (2017)	To evaluate how different types of Twitter users can influence the diffusion of famous brand content via retweets.	The results indicated that DI and information brokers are associated with a greater number of brand content retweets. The diffusion of these contents by information brokers is amplified when the original tweets mention DI.
Lin, Bruning and Swarna (2018)	To describe how online influencers can be used as tools for promoting products and services.	The main result is the proposition of a planning process in 5 stages, which can serve as a guide for entrepreneurs/brands to choose the ideal endorsers for their products and services.
Hughes, Swaminathan and Brooks (2019)	To study the factors that promote the success of a brand's online engagement based on DI marketing and sponsored blogging.	The results highlight the key role of interaction between the following factors: type of digital platform used, advertising campaign intent, source and type of content.

Source: Prepared by the authors (2020).

Kim and Kim (2018) examined the effect of sharing and social influence on Facebook and Twitter on sales. The authors used linear regression and found that consumer-generated social referrals to businesses significantly increase sales in social commerce. Similarly, Tien, Rivas and Liao (2019) studied the impact of the influence of DI on the purchase decision and, through structural equations, demonstrated that persuasion and perceived information and knowledge of the source are adequate predictors of influence of digital influencers.

Djafavora and Rushworth (2017) investigated Instagram's impact on source credibility, consumer purchase intent, and social identification with different types of celebrities. The authors conducted in-depth interviews and the survey results showed that Instagram celebrities influence the buying behavior of young users; however, non-traditional celebrities (digital influencers) are more powerful as survey participants perceive DIs for greater credibility and identification.

Hwang and Zhang (2018) used structural equation modeling to explore how parasocial relationships and electronic word-of-mouth (eWOM) affect followers' buying decisions. The results showed that the parasocial relationship moderates the paths between the followers'

knowledge of persuasion and the purchase and eWOM intentions. Complementarily, Sokolova and Kefi (2020) assessed how blogger credibility and parasocial interaction (PSI) are related and influence purchase intention and, through regression models, concluded that parasocial interaction (PSI) exhibits significant and positive relationships with the purchase intention.

Casaló, Flavián and Ibáñez-Sánchez (2018) identified some antecedents and main consequences of opinion leadership in the fashion industry, by using regression models. The authors concluded that aspects such as originality and uniqueness positively influence opinion leadership on Instagram; opinion leadership affects the consumer's intention to follow the influencer's advice; opinion leadership affects influencer-related behavioral intentions; and perceived fit and propensity for online interaction act as moderators.

Applying structural equation modeling, Alalwan (2018) concluded that performance expectation, hedonic motivation, interactivity, informativeness and perceived relevance have a significant impact on a customer's purchase intention. Jiménez-Castillo and Sánchez-Fernández (2019), in turn, from a sample of 280 followers, concluded that the perceived influential power of digital influencers not only helps generate engagement, but also increases expected value and behavioral intent regarding recommended brands. Furthermore, brand engagement in self-concept increases the expected value of the brand and the two variables also affect the intention to buy recommended brands.

Also on this topic, Dhanesh and Duthler (2019) examined the effects of awareness of endorsements paid to social media influencers on followers' knowledge of cognitive persuasion (ad recognition), knowledge of attitudinal persuasion (influencer relationships) and behavioral intentions, specifically eWOM intent and purchase intent. The authors used correlation and linear regression models and concluded that: paid endorsement awareness is correlated with ad recognition and the influencer-follower relationship, which are correlated with purchase intentions and eWOM. However, the recognition of ads does not affect the influencer-follower relationship; disclosure of paid endorsements improves followers' perceptions of influencer openness and transparency; the influencer's perceptions of openness and transparency can positively affect influencer-follower relationships and behavioral outcomes.

Arora et al. (2019) proposed a mechanism to measure the influence rate of IDs on popular social media platforms (Facebook, Twitter and Instagram). The results indicated that engagement, disclosure, sentiment and growth play a key role in determining influencers.

3 Methodology

3.1 Data

To achieve the objectives of this study, time series of daily sales of 10 luxury companies in the beauty and cosmetics retail industry of the second largest department store in the United Kingdom were considered. The official profiles of the brands on Instagram were analyzed and a discrepancy was noticed in the numbers presented, as the brand with the highest number of followers on that social media has about 30 million users and the brand with the lowest number of followers has about 50 thousand users. The size and commercial trajectory are also diverse, as the sample includes traditional brands and brands that were recently launched in the market. As for digital influencers, those that necessarily appear in the brand's official profile were considered in this study. Most DIs generate content in the area of beauty and cosmetics and have a number of followers in excess of 10,000 users. The fact that the number of followers is not that expressive is perhaps justified by the fact that they are luxury brands and, therefore, the popularity of influencers is not a decisive hiring requirement. Thus, information from posts on the brands' official profiles on Instagram and the publications made by hired digital influencers who tagged the brand's official profile were collected. The period analyzed was 2017-2019.

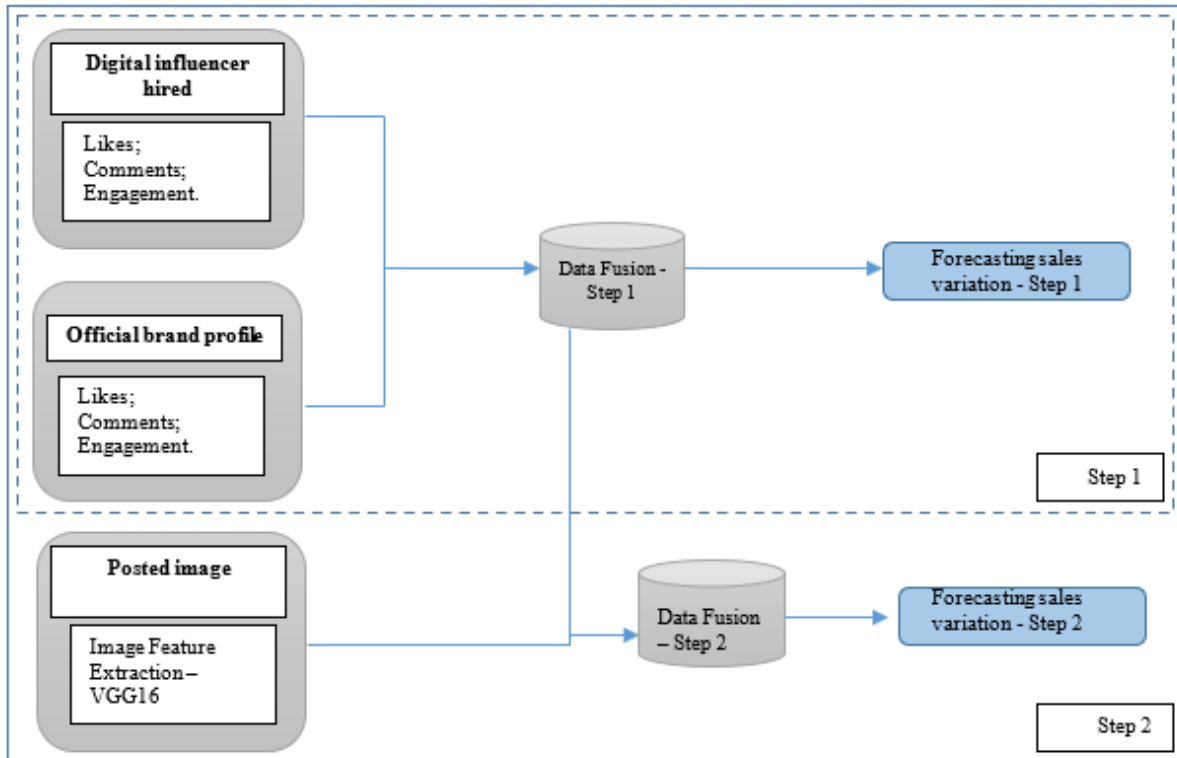
Instagram was the social media chosen because, according to Highfield and Leaver (2014), it is more focused on the commercial aspect through efficient data management and interaction between users, which enables advertising with a high level of segmentation, increasing the potential generator of electronic word of mouth (eWOM) and the interaction with brands (De Veirman, Cauberghe, & Hudders, 2017). In addition, several studies emphasize the importance of including Instagram in academic studies, as the network makes it possible to obtain information related to social, cultural and environmental issues about people's activities through quantitative and qualitative analyses of user behavior, (Hu, Manikonda, & Kambhampati, 2014; Lahuerta-Otero & Cordero-Gutiérrez, 2016).

To assess the importance of social media, specifically posts made by digital influencers on Instagram, univariate prediction models were augmented with variables related to posts.

The variables related to Instagram are subdivided into two specific groups: the first group of variables is composed of numerical data from the official profile of the company studied, and the post made by the hired digital influencer (likes, comments and engagement); the second group is basically composed of the image posted by the digital influencer (Figure 1).

Figure 1

Research Model – Forecast Stages 1 and 2



Source: Prepared by the authors (2020).

In view of this categorization of variables, data analysis was divided into two stages, namely (Figure 1):

- Step 1: development of forecast models, considering the historical behavior of daily variations in sales revenue of the ten companies studied, with the inclusion of numerical variables related to posts by digital influencers and the official profile of the company (variable group 1):

- Step 2: development of forecast models based on the historical behavior of daily variations in sales revenue of the ten companies studied considering the fusion of numerical data from Instagram posts and the extraction of variables related to the characteristics of the images posted by digital influencers (group of variables 1 and 2).

3.2 Applied methods

The core of artificial intelligence, models based on machine learning have been increasingly used in different fields of knowledge. In the marketing area, for example, Lee et al. (2018) and Timoshenko and Hauser (2019) adopted techniques such as SVM (Support Vector Machines) and Deep Neural Networks to assess consumer engagement and needs via

social media, highlighting the superior performance of machine learning models over other traditional methods. Also using data from social media, Cui, Gallino, Moreno and Zhang (2018) demonstrated the usefulness of machine learning models in predicting a company's sales. One of the reasons for this growing importance and use of machine learning models is precisely the power they have shown to learn from large volumes of data and identify patterns (Jordan & Mitchell, 2015).

In line with recent literature, the forecast of the time series of daily sales variations of the 10 companies selected in this work was made using the following methods, based on machine learning: Support Vector Machines (SVM) (Suykens & Vandewalle, 1999), Random Forest (Liaw & Wiener, 2002), XGBoost (Chen, He, Benesty, Khotilovich & Tang, 2015) and the method based on Deep Neural Networks Long Short-Term Memory, LSTM (Simonyan & Zisserman, 2014).

The SVM method uses the concept of decision plans to perform data separation in the sample space, through a kernel function (Suykens & Vandewalle, 1999). Hence, the SVM creates different geometric groups and recognizes different patterns in the data space. After the geometric definition of the data space and pattern recognition, it is possible to generate inferences and make predictions (Suykens & Vandewalle, 1999). The more diverse the SVM model training set, the better the inference results.

The Random Forest method is composed of a set of decision trees (Liaw & Wiener, 2002). In this method, data is divided into decision tables in which, as decisions are made, the algorithm walks between the nodes of the trees until it reaches the leaf. Then, the final inference is made. This method stands out for its simplicity and interpretability, as the end user can see the reason for each decision made by the algorithm (Liaw & Wiener, 2002).

XGBoost is a variation of the improved decision tree algorithm, through the gradient-descending method (Chen et al., 2015), which has a high hit rate on structured data due to its structure based on Gradient-boosting.

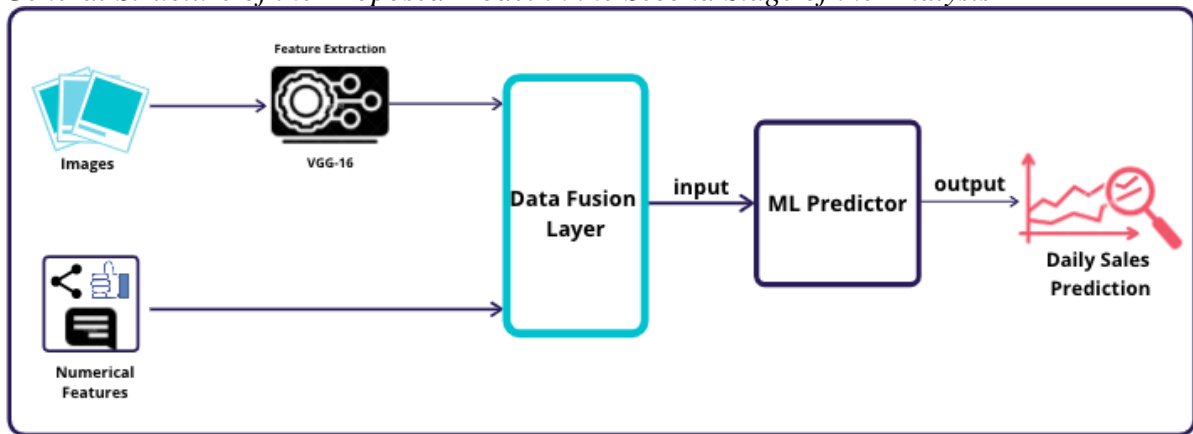
LSTM is a deep learning neural network or Deep Learning structured for the prediction task. It is composed of a recurrent structure in which information persists over time (temporal component) (Simonyan & Zisserman, 2014). The LSTM network is able to learn temporal dependencies and has a forgetting factor that is activated when patterns are not used for a period. The LSTM network has benefited from increased computational complexity to add a greater number of neurons, enabling more accurate learning (Simonyan & Zisserman, 2014).

Furthermore, its structure allows for parallel work with large data sets, such as those proposed in this research.

For the extraction of characteristics from the images of the posts, the VGG-16 Convolutional Neural Network was used. According to Simonyan and Zisserman (2014), the VGG-16 network is a state-of-the-art architecture for image feature extraction. In this work, activation vectors (1X4096) of the fully connected layer of the VGG-16 network were selected as input to the models, as illustrated in Figure 2.

Figure 2

General Structure of the Proposed Model in the Second Stage of the Analysis



Source: Prepared by the authors (2020).

The data fusion layer is responsible for the union of numerical data with data from images (posts), which are used as inputs to feed machine learning models. To evaluate the proposed models, the Root Mean Square Error (RMSE) metric given by Equation (1) was used:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (1)$$

where, n is the number of samples, \hat{y} is the predicted value. y is the real value of the series, and i is the time index of the time series. Cross-validation was used to measure the precision error of the forecasts of the daily sales variation time series, from the selection of specific cutoff points in the data history. The RMSE is an interesting metric in this context, as it measures the distance between the values predicted by the model and the actual values of the

series, penalizing large distances due to the quadratic difference. In other words, the lower the RMSE, the better the model.

Furthermore, to test whether the statistical differences between the fitted models were significant, an analysis of variance (ANOVA) test was implemented based on the RMSEs obtained.

4 Analysis of results

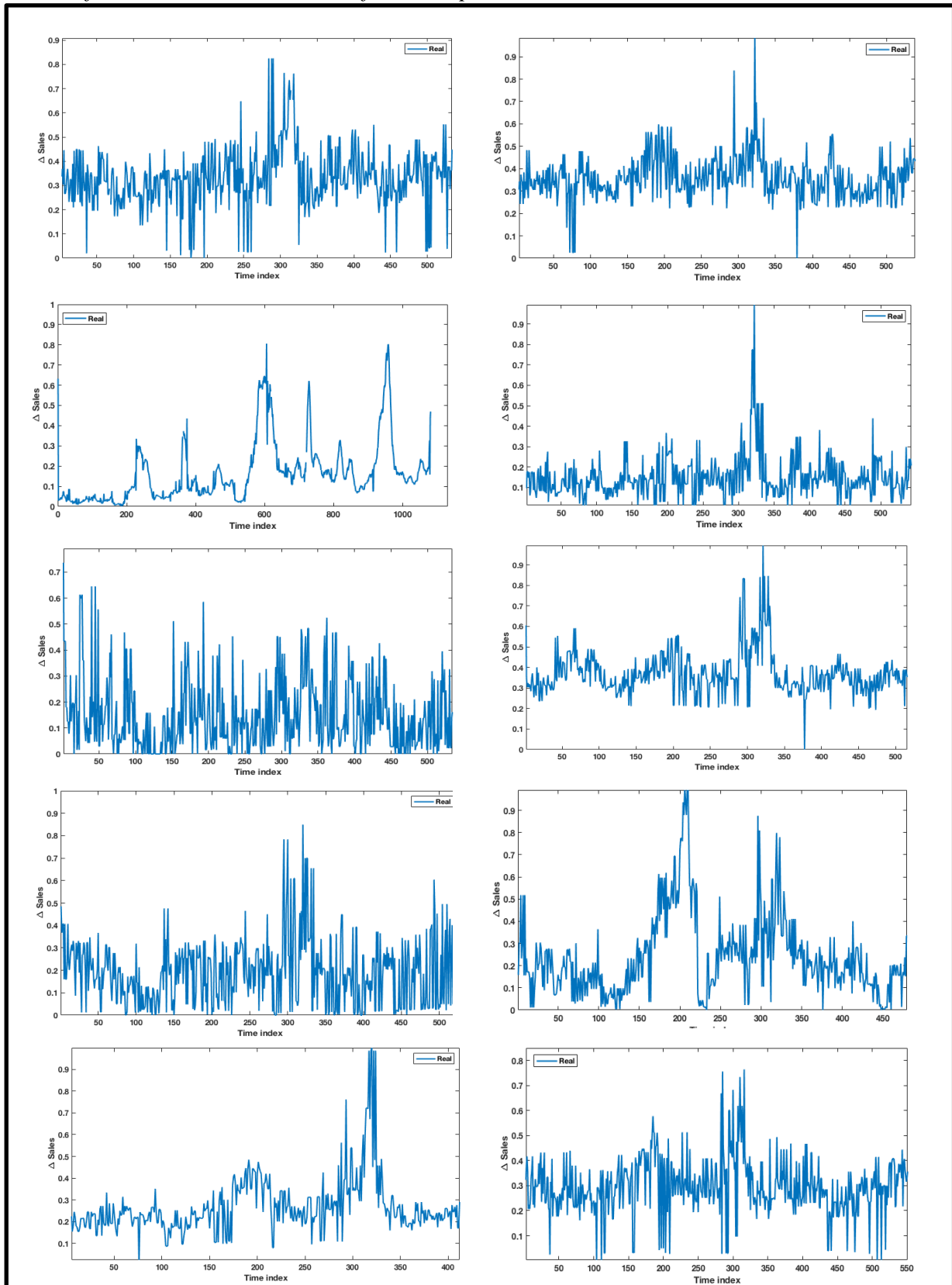
To answer the guiding question of this work and evaluate the predictive power of variables related to posts made by digital influencers (Likes, comments, engagement and images), the results for the steps of analysis proposed in this study are presented in this section: i) forecast of daily sales revenue variation based on numerical variables from Instagram posts; ii) addition of images (posts) to the forecasts from the data fusion. The prediction models used in this work contemplate the state of the art of machine learning models.

Initially, the original series of sales revenue variations of the 10 companies studied are exposed (Figure 3). The variation of sales revenue was adopted to smooth the data variability.

From Figure 3, it is possible to observe the historical behavior of the daily sales variations of the studied companies. The high variability of the data can be explained by the seasonal nature of the marketing strategies adopted by companies to attract and retain customers. The launch of new products and campaigns that prioritize the consumer experience in the physical store summarizes the strategies adopted by companies in recent years.

Figure 3

Series of Sales Revenue Variation of the Companies Studied



Source: Prepared by the authors (2020).

The interventionist character of advertising campaigns, especially through digital influencers, is quantitatively demonstrated when these events are configured as responsible for the structural break or for the change in behavior in the historical series of sales. In this study, the forecasting models of daily sales variations with the inclusion of variables related to the marketing campaigns of digital influencers will be able to assess the interventionist character of such strategies.

4.1 Step 1 - Forecasts of variation in daily sales considering numerical variables

To comply with the first analysis stage proposed in this work, the series of daily sales variations were used together with the numerical characteristics of the posts of digital influencers and the official profile of the company (group of variables 1). The results of forecasts of daily sales variations for the companies considered in this study are presented in Table 1.

Table 1

RMSE of the Forecast Models Used

Models	Emp 1	Emp 2	Emp 3	Emp 4	Emp 5	Emp 6	Emp 7	Emp 8	Emp 9	Emp 10
LSTM	0.0504	0.0798	0.0517	0.1065	0.0898	0.0256	0.1374	0.0249	0.1137	0.0617
SVM	0.0641	0.1192	0.0582	0.1037	0.0906	0.1173	0.1247	0.0856	0.1201	0.0674
R. Forest	0.0689	0.1641	0.0631	0.1545	0.1238	0.0560	0.1735	0.0622	0.1756	0.0767
XGBoost	0.0945	0.1072	0.1329	0.1031	0.0998	0.1987	0.0994	0.1262	0.1436	0.1023

Source: Prepared by the authors (2020).

From Table 1, it is possible to infer that the adjusted models presented a reasonable performance in predicting the sales variation of the companies analyzed. However, the method based on deep learning, LSTM, showed better results in 8 of the 10 series used. Thus, the results demonstrate the possibility of predicting daily sales based on data from social media. In this case, it is possible to affirm that there is a relationship between business strategies on Instagram, especially marketing campaigns through digital influencers, and the sales performance of the companies analyzed.

The results shown here agree with the main findings in the literature, showing the superior performance of machine learning models and the impacts of marketing via digital influencers on consumer behavior. In particular, the results are in line with what was presented by Cui et al. (2018). The authors obtained an increase in accuracy in predicting a company's daily sales through the use of machine learning models in conjunction with data from Facebook. Furthermore, the explanatory power of these types of data on consumer behavior was also corroborated by Lee et al. (2018) and Lou and Yuan (2019).

It is also noteworthy that the results found are consistent with the empirical evidence presented by Kim and Kim (2018). These authors argue that the social references generated by digital influencers are important aspects that affect the consumer's purchase decision. In this sense, there is also consistency with the study conducted by Djafavora and Rushworth (2017) since the authors signaled the importance of social identifications in the purchase decision process.

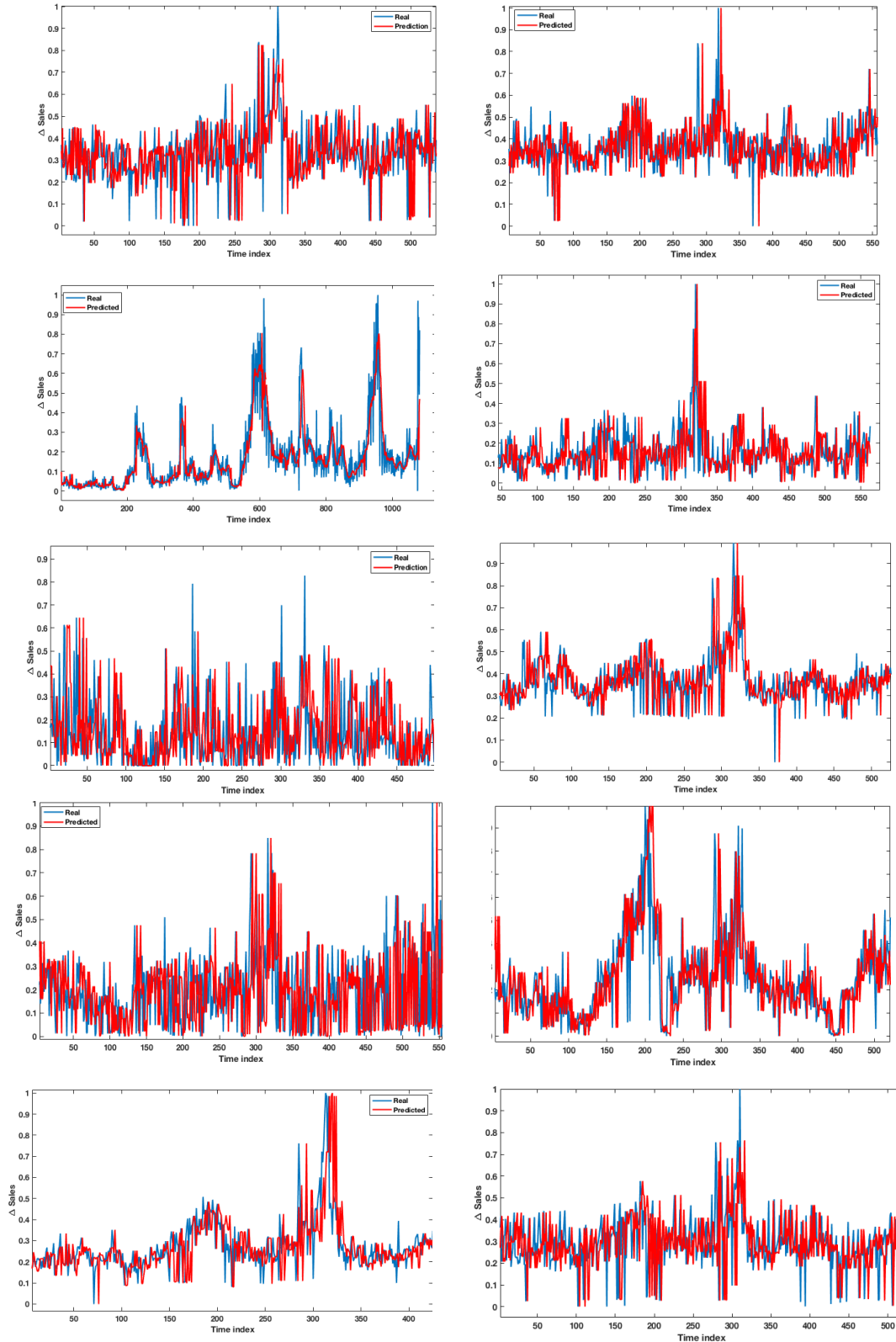
Figure 4 shows the time series of daily sales variations (blue line) and the time series predicted by the LSTM method (red line). It is possible to notice, through the figures, the degree of assertiveness of the predictions generated by the deep neural network model.

In addition to the possibility of using Instagram data to forecast sales variations, the adjusted statistical models allow for the identification of the variables that have mostly contributed to the degree of accuracy of the forecasts generated by the models. Through the importance of the variables provided by the XGBoost model, it is possible to notice that the factor that have mostly influenced the forecasts was the Likes in posts in which the company was tagged by the digital influencer. This factor was considered the most important for eight of the ten companies in the study. This finding reinforces the idea that the identification and social reference created by the digital influencer are configured as an efficient strategy to influence the consumer's purchase decision (Tien, Rivas, & Liao, 2019).

For two of the companies studied, the engagement rate of the digital influencer proved to be the most important variable during the forecasting process. This result is in line with the study conducted by Arora et al. (2019), in which the authors showed that the engagement variable was one of the most explanatory in determining the digital influencer. In general terms, it is observed that the most important variables in understanding and forecasting the historical sales pattern of the cases analyzed are not related to the official profile of the companies. The most relevant variables in the forecast models belong to the profile of the digital influencer hired (likes and engagement).

Figure 4

Forecast of the Time Series for the Companies Considered in The Study Based on Numerical Data From Social Networks



Source: Prepared by the authors (2020).

Thus, from the cases studied, the relevance of the use of digital influencers in marketing and advertising campaigns and the importance of the content generated by them in social media is highlighted, since it was demonstrated that DIs can generate and/or increase the public's connection with the brand, maintaining an open communication channel with followers (Childers, Lemon, & Hoy, 2019).

4.2 Step 2 - daily sales variation forecasts considering data fusion

In the second stage of analysis, we proceeded to merge the numerical data from Instagram (variables inserted in the first stage of analysis) with the images posted by the digital influencers, as shown in Figures 1 and 2. The purpose of this analysis stage is to verify that the characteristics extracted from the images posted through the VGG-16 Convolutional Neural Network contribute to the understanding of the historical behavior of the sales variations of the companies studied.

In general terms, it is believed that the recognition of feature patterns of posted images has the potential to improve the accuracy of prediction models as it extracts and provides the learning algorithm with unique information about the images and posts that cannot easily be found in the form of data. For this, the activation vectors (1X4096) of the fully connected layer of the VGG-16 network were used as input to the models along with the numerical variables. The results of this stage of the study are shown in Table 2.

Table 2

RMSE of the forecast models used

Models	Emp 1	Emp 2	Emp 3	Emp 4	Emp 5	Emp 6	Emp 7	Emp 8	Emp 9	Emp 10
LSTM	0.0422	0.0615	0.0435	0.0898	0.0682	0.0512	0.0424	0.0382	0.0492	0.0430
SVM	0.0728	0.0985	0.0500	0.1002	0.0859	0.1145	0.2894	0.0631	0.0579	0.0954
R. Forest	0.0496	0.0635	0.0482	0.0976	0.0752	0.0745	0.1735	0.0622	0.1756	0.0767
XGBoost	0.0528	0.1285	0.0463	0.1421	0.0933	0.0721	0.0528	0.0441	0.0533	0.0521

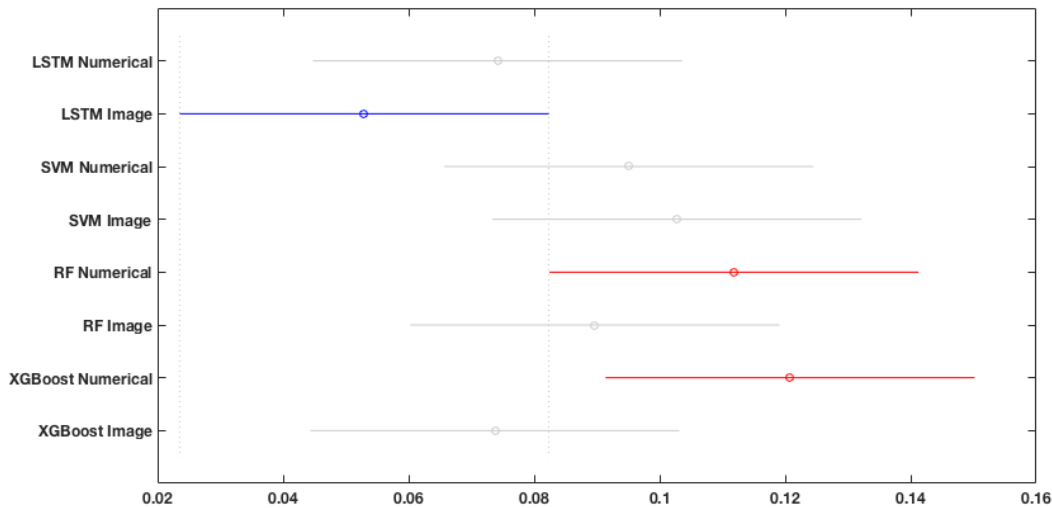
Source: Prepared by the authors (2020).

From the information shown in Table 2, it is possible to notice that the method based on deep learning (LSTM) obtained the best results in terms of prediction accuracy when compared to other traditional methods of machine learning. The result was consensual for all the companies studied and this fact can be explained by the high learning capacity of the method, which is obtained through the use of millions of neurons, which are able to approximate the predicted values to the real values of a time series more efficiently.

Furthermore, it is possible to note that, in general, the extraction and addition of the characteristics of the images of each post improved, by merging the variables, the forecasts predicted for the daily sales variations of the companies. Thus, it is demonstrated that the image is an important communication channel that has a potential impact on the sales of the companies studied. This result is supported by the empirical findings of Bakhshi, Shamma and Gilbert (2014). The authors demonstrated that the presence of images on social media increases the quantity of likes and comments on the post, and user engagement and, consequently, the influence on the consumer’s purchase decision process increases.

Figure 5

ANOVA Test to Analyze the Statistical Difference Between the Adjusted Models



Source: Prepared by the authors (2020).

The ANOVA test, shown in Figure 5, was performed for all prediction models adjusted in both step 1 (numerical data from posts) and step 2 (fusion of numeric data and images). Through the results presented, the statistical difference of the forecast models adjusted in step 2 was evidenced compared to the forecast models adjusted in step 1.

The ANOVA test was used to analyze, in general terms, the performance of the adjusted forecasting models and to verify the existence of statistical differences between the different methods and combinations of variables tested for all companies. Thus, this result indicates the efficiency of using post images as an input variable to predict variations in daily sales based on data from social networks, specifically Instagram.

5. Final considerations

Undoubtedly, social media has become an effective means of promoting marketing campaigns and connecting consumers and brands. In this study, the objective was to study the relevance of Instagram posts in the construction of forecasting models for daily sales revenue variation. For this, the daily sales of ten retail companies in the beauty and cosmetics industry were analyzed.

In the first part of the study, the predictive character of variables from Instagram was tested. The series of daily sales variations of the ten companies were modeled considering the characteristics of posts by digital influencers and the official profile of the company. The models used proved to be efficient in forecasting and, through XGBoost, it was possible to observe that variables related to the profile of the digital influencer are preponderant for the quality of the forecast. The importance of likes and engagement variables reinforces the idea that the identification and social reference generated by the DI are important aspects in the purchase decision process. The results of this step are consistent with the evidence presented by Cui et al. (2018), Lee et al. (2018) and Lou and Yuan (2019).

In the second stage of the study, an innovative approach was proposed to merge numerical data with post image data for the sales series forecasting process. From the adjustment of the models, it was concluded that the images are responsible for adding unique attributes that help in predicting and understanding the patterns of the analyzed sales series. It can be said that machine learning models incorporated both numerical variables and captured image patterns to generate more accurate predictions. The results of this stage of the analysis were consistent with the empirical results found by Bakhshi, Shamma and Gilbert (2014) and

extended the results found by the authors, when using and proposing a methodology for extracting and quantifying the characteristics of an image.

Overall, promisingly, the study showed the efficiency of methods based on machine learning in predicting daily sales variation from Instagram data, especially with regard to the incorporation and extraction of image data (quantitative visual analysis).

We emphasize the importance of studies of this nature for the formulation of more efficient marketing strategies, which allow for both the identification of digital influencers with greater potential for influence and results, and the strategic formulation of posts that have a greater impact on the public and conversion in sales results through the use and extraction of information from image analysis. The combination of these factors could generate the maximization of influence and favor the expansion of viral marketing through online social networks more efficiently (More & Lingam, 2017).

The limitations of this work are related to the limited number of companies included in the study and the use of posts made only on Instagram to study the increase in the predictive power of the adjusted statistical models. Therefore, for future work, the use of other social media such as Twitter and Facebook is suggested for the purpose of comparing forecast models. Furthermore, the development of studies of this nature for other industries would be of great importance for the generation of new empirical evidence. The inclusion of the content of the post comments, through the analysis of feelings, as one of the analysis variables for the prediction models, is also suggested. The combination of methods is believed to enrich the process and increase the degree of accuracy of the developed models. Finally, a comparison between the performance of digital influencers and traditional professional advertising is suggested.

References

- Abolghasemi, M., Eshragh, A., Hurley, J., & Fahimnia, B. (2019). Demand Forecasting in the Presence of Systematic Events: Cases in Capturing Sales Promotions. arXiv preprint arXiv:1909.02716.
- Agnihotri, R., Dingus, R., Hu, M. Y., & Krush, M. T. (2016). Social media: Influencing customer satisfaction in B2B sales. *Industrial Marketing Management*, 53, 172-180. <https://doi.org/10.1016/j.indmarman.2015.09.003>
- Alalwan, A. A. (2018). Investigating the impact of social media advertising features on customer purchase intention. *International Journal of Information Management*, 42, 65-77. <https://doi.org/10.1016/j.ijinfomgt.2018.06.001>

- Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177-1190. <https://doi.org/10.1016/j.tele.2017.05.008>
- Araujo, T., Neijens, P., & Vliegenthart, R. (2017). Getting the word out on Twitter: The role of influentials, information brokers and strong ties in building word-of-mouth for brands. *International Journal of Advertising*, 36(3), 496-513. <https://doi.org/10.1080/02650487.2016.1173765>
- Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index-insights from Facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, 49, 86-101. <https://doi.org/10.1016/j.jretconser.2019.03.012>
- Asur, S., & Huberman, B.A. (2010). "Predicting the Future with Social Media", available at <http://www.hpl.hp.com/techreports/2010/HPL-2010-53.pdf>
- Audrezet, A., De Kerviler, G., & Moulard, J. G. (2018). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of Business Research*, 177, 557-569. <https://doi.org/10.1016/j.jbusres.2018.07.008>
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318. DOI: <https://doi.org/10.1509/jmr.14.0380>
- Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: photos with faces attract more likes and comments on Instagram. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '14*. Association for Computing Machinery, New York, NY, USA, 965–974. <https://doi.org/10.1145/2556288.2557403>
- Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2018). Influencers on Instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*, 177, 510-519. DOI: <https://doi.org/10.1016/j.jbusres.2018.07.005>
- Chen, I. F., & Lu, C. J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, 28(9), 2633-2647. DOI: <https://doi.org/10.1007/s00521-016-2215-x>
- Childers, C. C., Lemon, L. L., & Hoy, M. G. (2019). # Sponsored# Ad: Agency perspective on influencer marketing campaigns. *Journal of Current Issues & Research in Advertising*, 40(3), 258-274. DOI: <https://doi.org/10.1080/10641734.2018.1521113>
- Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1-4.
- Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management*, 27(10), 1749-1769. <https://doi.org/10.1111/poms.12707>

- De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. *International journal of advertising*, 36(5), 798-828. <https://doi.org/10.1080/02650487.2017.1348035>
- De Vries, N. J., & Carlson, J. (2014). Examining the drivers and brand performance implications of customer engagement with brands in the social media environment. *Journal of Brand Management*, 21(6), 495-515. <https://doi.org/10.1057/bm.2014.18>
- Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, 68, 1-7. <https://doi.org/10.1016/j.chb.2016.11.009>
- Dhanesh, G. S., & Duthler, G. (2019). Relationship management through social media influencers: Effects of followers' awareness of paid endorsement. *Public Relations Review*, 45(3), 101765. <https://doi.org/10.1016/j.pubrev.2019.03.002>
- Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*, 61, 47-55. <https://doi.org/10.1016/j.chb.2016.03.003>
- Felix, R., Rauschnabel, P. A., & Hinsch, C. (2017). Elements of strategic social media marketing: A holistic framework. *Journal of Business Research*, 70, 118-126. <https://doi.org/10.1016/j.jbusres.2016.05.001>
- Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570-576. <http://www.jstor.org/stable/20141547>
- Fildes, R., Goodwin, P., & Önkal, D. (2019). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, 35(1), 144-156. <https://doi.org/10.1016/j.ijforecast.2017.12.006>
- Fuchs, C. (2017). *Social media: A critical introduction*. (2nd ed.) London: Sage.
- Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of interactive marketing*, 27(4), 242-256. <https://doi.org/10.1016/j.intmar.2013.09.004>
- Gillon, K., Aral, S., Lin, C. Y., Mithas, S., & Zozulia, M. (2014). Business analytics: radical shift or incremental change? *Communications of the Association for Information Systems*, 34(1), 13. <https://doi.org/10.17705/1CAIS.03413>
- Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R., & Singh, R. (2016). Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. *Journal of business research*, 69(12), 5833-5841. <https://doi.org/10.1016/j.jbusres.2016.04.181>

- Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, 30(2), 127-135. [https://doi.org/10.1016/S0305-0483\(01\)00062-7](https://doi.org/10.1016/S0305-0483(01)00062-7)
- Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). The predictive power of online chatter. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 78-87). <https://doi.org/10.1145/1081870.1081883>
- Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, 56(3), 387-404. <https://doi.org/10.2501/IJMR-2014-025>
- Hajli, N., & Sims, J. (2015). Social commerce: The transfer of power from sellers to buyers. *Technological Forecasting and Social Change*, 94, 350-358. <https://doi.org/10.1016/j.techfore.2015.01.012>
- Highfield, T., Leaver, T. (2014). A methodology for mapping Instagram hashtags. *First Monday*, 20(1). DOI: <https://doi.org/10.5210/fm.v20i1.5563>
- Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What We Instagram: A First Analysis of Instagram Photo Content and User Types. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1). Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/14578>.
- Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2), 738-748. <https://doi.org/10.1016/j.ejor.2014.02.022>
- Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. (2016). The influence of social media interactions on consumer-brand relationships: A three-country study of brand perceptions and marketing behaviors. *International Journal of Research in Marketing*, 33(1), 27-41. <https://doi.org/10.1016/j.ijresmar.2015.06.004>
- Hughes, C., Swaminathan, V., & Brooks, G. (2019). Driving brand engagement through online social influencers: An empirical investigation of sponsored blogging campaigns. *Journal of Marketing*, 83(5), 78-96. <https://doi.org/10.1177/0022242919854374>
- Hwang, K., & Zhang, Q. (2018). Influence of parasocial relationship between digital celebrities and their followers on followers' purchase and electronic word-of-mouth intentions, and persuasion knowledge. *Computers in Human Behavior*, 87, 155-173. <https://doi.org/10.1016/j.chb.2018.05.029>
- Hyndman, R. J., & Athanasopoulos, G. (2014). Optimally reconciling forecasts in a hierarchy. *Foresight: The International Journal of Applied Forecasting*, (35), 42-48.
- Jiménez-Castillo, D., & Sánchez-Fernández, R. (2019). The role of digital influencers in brand recommendation: Examining their impact on engagement, expected value and purchase intention. *International Journal of Information Management*, 49, 366-376. <https://doi.org/10.1016/j.ijinfomgt.2019.07.009>

- Jin, S. V. (2018). "Celebrity 2.0 and beyond!" Effects of Facebook profile sources on social networking advertising. *Computers in Human Behavior*, 79, 154-168. DOI: <https://doi.org/10.1016/j.chb.2017.10.033>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- Kim, N., & Kim, W. (2018). Do your social media lead you to make social deal purchases? Consumer-generated social referrals for sales via social commerce. *International Journal of Information Management*, 39, 38-48. <https://doi.org/10.1016/j.ijinfomgt.2017.10.006>
- Kim, E., & Kim, Y.-C. (2018). Communication Infrastructure, Migrant Community Engagement, and Integrative Adaptation of Korean Chinese Migrants in Seoul. *Communication Research*, 1-23. <https://doi.org/10.1177/0093650218782819>
- Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*, 181, 145-153. DOI: 10.1016/j.ijpe.2015.09.011
- Kremer, M., Siemsen, E., & Thomas, D. J. (2016). The sum and its parts: Judgmental hierarchical forecasting. *Management Science*, 62(9), 2745-2764. <https://doi.org/10.1287/mnsc.2015.2259>
- Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing*, 80(1), 7-25. <https://doi.org/10.1509/jm.14.0249>
- Kulkarni, G., Kannan, P. K., & Moe, W. (2012). Using Online Search Data to Forecast New Product Sales. *Decision Support Systems*, 52 (3), 604-611. <https://doi.org/10.1016/j.dss.2011.10.017>
- Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, 64, 575-583. <https://doi.org/10.1016/j.chb.2016.07.035>
- Lamond, D., Dwyer, R., Ramanathan, R., Black, A., Nath, P., & Muyldermans, L. (2010). Impact of environmental regulations on innovation and performance in the UK industrial sector. *Management Decision*, 48(10), 1493-1513. <https://doi.org/10.1108/00251741011090298>
- Lassen, N. B., Madsen, R., & Vatrapu, R. (2014). Predicting iphone sales from iphone tweets. *2014 IEEE 18th International Enterprise Distributed Object Computing Conference* (pp. 81-90). IEEE. <https://doi.org/10.1109/EDOC.2014.20>
- Lawrence, M., Goodwin, P., O'Connor, M., & Önköl, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of forecasting*, 22(3), 493-518. <https://doi.org/10.1016/j.ijforecast.2006.03.007>

- Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, 64(11), 5105-5131. <https://doi.org/10.1287/mnsc.2017.2902>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R news*, 2(3), 18-22. Retrieved from: <https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf>
- Lin, H. C., Bruning, P. F., & Swarna, H. (2018). Using online opinion leaders to promote the hedonic and utilitarian value of products and services. *Business Horizons*, 61(3), 431-442. <https://doi.org/10.1016/j.bushor.2018.01.010>
- Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. *International Journal of Information Management*, 35(4), 490-503. <https://doi.org/10.1016/j.ijinfomgt.2015.04.001>
- Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70(3), 74-89. <https://doi.org/10.1509/jmkg.70.3.074>
- Lou, C., & Yuan, S. (2019). Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, 19(1), 58-73. <https://doi.org/10.1080/15252019.2018.1533501>
- Lu, B., Fan, W., & Zhou, M. (2016). Social presence, trust, and social commerce purchase intention: An empirical research. *Computers in Human Behavior*, 56, 225-237. <https://doi.org/10.1016/j.chb.2015.11.057>
- More, J. S., & Lingam, C. (2019). A SI model for social media influencer maximization. *Applied Computing and Informatics*, 15(2), 102-108. <https://doi.org/10.1016/j.aci.2017.11.001>
- Oliva, R., & Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and operations Management*, 18(2), 138-151. <https://doi.org/10.1111/j.1937-5956.2009.01003.x>
- Ramanathan, U., & Muyltermans, L. (2011). Identifying the underlying structure of demand during promotions: A structural equation modelling approach. *Expert systems with applications*, 38(5), 5544-5552. <https://doi.org/10.1016/j.eswa.2010.10.082>
- Schivinski, B., & Dabrowski, D. (2016). The effect of social media communication on consumer perceptions of brands. *Journal of Marketing Communications*, 22(2), 189-214. <https://doi.org/10.1080/13527266.2013.871323>
- Schouten, A. P., Janssen, L., & Verspaget, M. (2020). Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit. *International journal of advertising*, 39(2), 258-281. <https://doi.org/10.1080/02650487.2019.1634898>
- Shiau, W. L., Dwivedi, Y. K., & Lai, H. H. (2018). Examining the core knowledge on facebook. *International Journal of Information Management*, 43, 52-63. <https://doi.org/10.1016/j.ijinfomgt.2018.06.006>

-
- Simonyan, K., & Zisserman, A. (2014). Very Deep convolutional networks for large-scale image recognition. *Computer Science*. arXiv preprint arXiv:1409.1556.
- Slack, N., Chambers, S., & Johnston, R. (2009). Administração da produção (Vol. 2). São Paulo: Atlas.
- Sokolova, K., & Kefi, H. (2020). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. *Journal of Retailing and Consumer Services*, 53, 101742. <https://doi.org/10.1016/j.jretconser.2019.01.011>
- Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300. DOI: 10.1023/A:1018628609742
- Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1), 1-26. <https://doi.org/10.1016/j.ejor.2015.11.010>
- Tien, D. H., Rivas, A. A. A., & Liao, Y. K. (2019). Examining the influence of customer-to-customer electronic word-of-mouth on purchase intention in social networking sites. *Asia Pacific Management Review*, 24(3), 238-249. <https://doi.org/10.1016/j.apmr.2018.06.003>
- Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, 66(2), 299-307. <https://doi.org/10.1057/jors.2013.174>
- 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>



APENAS UMA POSTAGEM? PREVISÕES DE VENDAS DIÁRIAS DE EMPRESAS VAREJISTAS DE BELEZA E COSMÉTICO A PARTIR DA INFLUÊNCIA DE MÍDIAS SOCIAIS

 **Gabriel Rodrigo Gomes Pessanha**

Doutor em Administração

Universidade Federal de Alfenas (UNIFAL) – Instituto de Ciências Sociais Aplicadas (ICSA)

Varginha, MG - Brasil

gabriel.pessanha@unifal-mg.edu.br

 **Eduardo Almeida Soares**

Mestre

Universidade de Lancaster

Reino Unido – Inglaterra.

e.almeidasoares@lancaster.ac.uk

Objetivo: Estudar a relevância das postagens no *Instagram* na construção de modelos de previsão de variação de receitas de vendas diárias para empresas varejistas do setor de beleza e cosméticos.

Metodologia: Foram consideradas séries temporais de vendas diárias entre os anos de 2017 e 2019 de 10 empresas varejistas do setor de beleza e cosméticos. Métodos baseados em aprendizagem de máquina foram empregados e os modelos de previsões foram incrementados com variáveis numéricas do perfil oficial da empresa, da postagem feita pelo influenciador digital contratado e as características das imagens postadas pelo influenciador digital foram incluídas nos modelos.

Relevância e Originalidade: O estudo é inovador, pois ultrapassa as reflexões qualitativas sobre a temática e traz evidências empíricas quanto aos impactos na acurácia da previsão a partir da inclusão de variáveis de mídias sociais. Apresentou-se uma estratégia de fusão de dados (numéricos e imagens) para a previsão de vendas diárias de empresas de varejo do setor de beleza e cosméticos.

Principais resultados: Os modelos se mostraram eficientes na previsão e a importância das variáveis *likes* e engajamento reforça a ideia de que a identificação e referência social gerada pelo ID são importantes aspectos no processo de decisão de compra. Constatou-se que as imagens são responsáveis por adicionar atributos exclusivos que ajudam na previsão e no entendimento dos padrões das séries de vendas.

Contribuições teóricas e metodológicas: O estudo demonstrou, de modo promissor, a eficiência dos métodos baseados em aprendizagem de máquina na previsão de vendas a partir de dados do *Instagram*, especialmente, no que se refere à incorporação e extração de dados de imagens.

Palavras-chave: Mídias sociais. Imagens. Inteligência artificial. Previsão de vendas. *Marketing* digital. Influenciador digital.

Como citar

American Psychological Association (APA)

Pessanha, G. R. G., & Soares, E. A. (2021, out./dez.). Apenas uma postagem? previsões de vendas diárias de empresas varejistas de beleza e cosmético a partir da influência de mídias sociais. *Revista Brasileira de Marketing – ReMarK*, 20(4), 258 -285. <https://doi.org/10.5585/remark.v20i4.17914>.

1 Introdução

Diante do atual ambiente concorrencial, o planejamento se torna uma tarefa importante para gerar vantagens competitivas no processo de decisão. Nesse contexto, a previsão da demanda é essencial para o planejamento organizacional, uma vez que possibilita a compreensão das influências de fatores internos e externos no comportamento da demanda e sua medição futura (Chen & Lu, 2017; Kremer, Siemsen, & Thomas, 2016; Trapero, Kourentzes, & Fildes, 2015). Assim, ela se torna uma etapa essencial para o desenvolvimento de estratégias relacionadas a estoques, mão de obra, compras, planos de produção, além de gerenciar a satisfação do cliente e as economias financeiras em toda a cadeia de suprimentos (Kremer et al., 2016; Fildes, Goodwin, & Önkal, 2019).

Inúmeras variáveis podem influenciar a dinâmica da demanda e, portanto, os modelos estatísticos que consistem apenas em dados históricos de vendas geralmente são insuficientes para produzir previsões precisas (Abolghasemi, Eshragh, Hurley, & Fahimnia, 2019). A incerteza torna o planejamento e o controle mais difíceis (Slack, Chambers, & Johnston, 2009; Hyndman & Athanasopoulos, 2014). De acordo com Lawrence, Goodwin, O'Connor e Önkal (2006), embora os dados históricos sejam importantes para a geração de previsões, muitas vezes são insuficientes para gerar preditores precisos, pois os modelos não têm a capacidade de atualizar aspectos contextuais e incorporar novas informações.

Eventos relativos à promoção de produtos e serviços são importantes e podem aumentar a precisão dos modelos de previsão. Trapero et al. (2015) constataram que as campanhas de *marketing* influenciam significativamente a demanda do mercado. Diversos estudos, demonstraram que a incorporação de campanhas promocionais de *marketing* aumenta a precisão das previsões de vendas (Fildes & Goodwin, 2007; Goodwin, 2002). Embora diversas pesquisas tenham analisado o impacto das campanhas de *marketing* nas vendas (Alalwan, 2018; Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016; Lamond et al., 2010; Ramanathan & Muyldermans, 2011; Oliva & Watson, 2009; Huang, Fildes, & Soopramanien, 2014; Kourentzes & Petropoulos, 2016) Abolghasemi, Eshragh, Hurley e Fahimnia (2019) afirmam que a tarefa de quantificar o impacto das promoções ainda é problemática para a prática de analistas e pesquisadores acadêmicos, especialmente, quando se trata de campanhas de *marketing* de mídia social.

As mídias sociais se tornaram uma fonte indispensável de informação e conteúdo (Lou & Yuan, 2019; Shiau, Dwivedi, & Lai, 2018). Além disso, tais plataformas mostraram-se eficazes na promoção de campanhas de *marketing* (Alalwan, 2018; De Vries & Carlson, 2014; Gensler, Völckner, Liu-Thompkins, & Wiertz, 2013). De acordo com Sokolova e Kefi (2020), além das publicações e *fan pages*, os usuários geram e publicam ativamente conteúdo multimídia, incluindo suas opiniões sobre marcas e produtos. Nesse novo processo, os influenciadores digitais criam conteúdo atraente para comercializar produtos ou serviços, visando um nicho específico do público.

Assim, o uso de influenciadores digitais para campanhas de publicidade e *marketing* está se demonstrando cada vez mais popular entre as empresas, pois, de acordo com Childers, Lemon e Hoy (2019), os influenciadores digitais conectam o público às marcas, mantendo um canal de comunicação aberto com os seguidores. Ademais, alguns estudos (Welbourne & Grant, 2016; Gillon, Aral, Lin, Mithas, & Zozulia, 2014; Lipizzi, Iandoli, & Marquez, 2015) demonstram que a publicidade de influenciadores digitais é mais eficiente que a publicidade profissional tradicional.

Outras pesquisas evidenciaram o valor das mídias sociais para melhorar modelos de previsão de variação de vendas (Liu, 2006; Lassen, Madsen, & Vatrapu, 2014; Kulkarni, Kannan, & Moe, 2012; Asur & Huberman, 2010; Gruhl, Guha, Kumar, Novak, & Tomkins, 2005). Todavia, apesar do esforço para entender quantitativamente a influência das plataformas digitais e de seus usuários na demanda de produtos, ainda não há consenso sobre o impacto das redes sociais e dos influenciadores digitais no aumento de vendas e lucratividade. Sendo assim, a hipótese geral deste trabalho está pautada no poder preditivo das postagens feitas pelos influenciadores digitais. Em outras palavras, acredita-se que as características dos *posts* (*Likes*, comentários, engajamento e imagens) possam se configurar como variáveis preditivas com o poder de incrementar a acurácia dos modelos tradicionais, baseados unicamente no comportamento histórico de vendas.

O objetivo desta pesquisa é estudar a relevância das postagens no *Instagram* na construção de modelos de previsão de variação de receitas de vendas diárias para empresas varejistas do setor de beleza e cosméticos.

Além da seção introdutória, o artigo está estruturado da seguinte forma: na seção 2, apresenta-se o contexto teórico no qual a pesquisa foi desenvolvida; na seção seguinte, descrevem-se a metodologia empregada nas análises e os procedimentos para a comparação de desempenho de previsão dos modelos utilizados; na seção 4, as análises e a discussão dos

resultados são expostas; as considerações finais e direções futuras de trabalho são apresentadas na seção 5; e, por fim, a seção 6 traz as referências bibliográficas utilizadas no decorrer deste trabalho.

2 Referencial teórico

2.1 Mídias sociais e marketing – estudos anteriores

Nos últimos anos, alguns estudos investigaram a relação entre a mídia social e o comportamento do consumidor (Alalwan, Rana, Dwivedi, & Algharabat, 2017; Hajli & Sims, 2015; Hajli, 2014). No intuito de levantar os estudos mais citados dos últimos 5 anos, foi realizada uma busca na base de dados da *SCOPUS* com as palavras-chave “mídias sociais”, “influenciadores digitais” e “vendas”. O Quadro 1 apresenta uma síntese com alguns trabalhos anteriores relacionados a mídias sociais. A relação entre influenciadores digitais (IDs) e decisões de compra tem sido alvo de muitas pesquisas acadêmicas.

Jin (2018) analisou os efeitos do *Facebook* na publicidade por meio da ANOVA, análise de mediação e *Bootstrapping*. O autor conduziu dois experimentos que confirmam a teoria da justificativa no endosso de marcas de celebridades baseadas no *Facebook*. Os resultados da pesquisa indicam que o endosso a terceiros influencia mais positivamente a credibilidade da fonte e a eficácia dos anúncios do que o autoendosso. Além disso, a identificação social e o envolvimento são medidores e moderadores significativos.

More e Lingam (2019) forneceram uma visão geral da identificação do influenciador e da maximização da influência. Utilizando um modelo epidêmico de infecção, os autores concluíram que, ao identificar os usuários influentes nas mídias sociais, diferentes estratégias de negócios podem ser planejadas e que a maximização da influência e a mineração de influência social juntas irão formar os componentes significativos para permitir amplo *marketing* viral por meio de redes sociais *on-line*.

Já Audrezet, De Kerviler e Moulard (2018), por meio de observações não participativas e entrevistas semiestruturadas, estudaram as estratégias que os influenciadores digitais (IDs) usam para manter sua autenticidade pessoal ao fazer parceria com marcas. A partir dos resultados, duas estratégias de gerenciamento de autenticidade emergiram da análise: autenticidade apaixonada e transparente. Por fim, os autores propuseram uma estrutura de quatro vias que fornece a primeira conceituação de como os IDs podem gerenciar a autenticidade para lidar com as tensões criadas pela intrusão da marca em seu conteúdo.

Quadro 1

Alguns Estudos Anteriores Relacionados à Temática

Autor(es)	Objetivo	Principais resultados
Babić Rosário, Sotgiu, De Valck e Bijmolt (2016)	Realizar uma meta-análise para mensurar o impacto do eWOM nas vendas de diferentes produtos.	Em média, o eWOM está positivamente correlacionado com as vendas, porém seu impacto varia conforme o tipo do produto, plataforma digital e métrica utilizada.
Kumar, Bezawada, Rishika, Janakiraman e Kannan (2016)	Estudar o efeito dos conteúdos gerados por marcas nas mídias sociais em três métricas chaves do consumidor (gastos, compras cruzadas e lucratividade do consumidor).	Aliado à propaganda televisiva e o <i>marketing</i> por <i>e-mail</i> , o FGC teve um efeito positivo no comportamento do consumidor.
Lu, Fan e Zhou (2016)	Teorizar a natureza do aspecto social no mercado de <i>social commerce</i> (SC) (comércio social?) <i>on-line</i> , propondo três variáveis de presença social.	Fatores de presença social baseados em tecnologias sociais têm um efeito positivo significativo no comércio <i>on-line</i> .
Erkan e Evans (2016)	Examinar o efeito do eWOM em mídias sociais na intenção de compra dos consumidores.	A qualidade, credibilidade, utilidade e adoção de informação, necessidade de informação e atitude com relação à marca são fatores chave do eWOM em mídias sociais que influenciam a intenção de compra dos consumidores.
De Veirman, Cauberghe e Hudders (2017)	Avaliar, com base no número de seguidores no <i>Instagram</i> , como influenciadores digitais são enxergados e como eles podem alterar a percepção de consumidores com relação a uma marca e à originalidade de um produto.	Influenciadores digitais (IDs) com um alto número de seguidores tendem a ser enxergados de uma forma melhor, mas nem sempre isso se converte em um maior poder de formação de opinião entre seus seguidores. Seguir poucos perfis, no entanto, pode prejudicar a imagem de um ID. Ademais, IDs muito populares tendem a prejudicar a percepção dos consumidores com relação à originalidade de um produto ou uma marca.
Godey et al. (2016)	Estudar o efeito do <i>marketing</i> via mídias sociais na criação de valor de marca e no comportamento de consumidores com respeito a marcas pioneiras no setor de bens de luxo.	O estudo demonstra as relações entre ações de <i>marketing</i> via mídias sociais e a preferência por uma marca, preços <i>premium</i> e fidelização. Outra contribuição dos autores é na descoberta de que as SMMEs têm um impacto positivo no valor da marca e suas duas principais dimensões são: percepção da marca e imagem da marca.
Schivinski e Dabrowski (2016)	Analisar os efeitos da comunicação via mídias sociais na percepção de consumidores a respeito de marcas	Comunicação em mídias sociais via usuário padrão influencia positivamente tanto o valor da marca quanto a atitude em relação a ela, ao passo que a comunicação via marca afeta apenas a atitude em relação à marca. Foi demonstrado também que o valor da marca e a atitude da marca têm uma influência positiva nas intenções de compras.

Autor(es)	Objetivo	Principais resultados
Hudson, Huang, Roth e Madden (2016)	Avaliar como o uso das mídias sociais influencia nas relações consumidor-marca.	O uso das mídias sociais está positivamente relacionado à qualidade da interação entre consumidor e marca. Esse efeito é intensificado com a personificação de marcas.
Agnihotri, Dingus, Hu e Krush (2016)	Testar os efeitos dos comportamentos de comunicação informativa de vendedores no uso das mídias sociais e a satisfação do consumidor.	Foi observado que o uso de mídias sociais por parte de vendedores tem um impacto nos comportamentos de comunicação informativa, intensificando a responsividade do vendedor e a satisfação do consumidor.
Lee, Hosanagar e Nair (2018)	Descrever o efeito da publicidade via mídias sociais no engajamento do consumidor por meio de dados do <i>Facebook</i> .	A inclusão de conteúdo relacionado à personalidade de uma marca, como humor e emoção, está associada a maiores níveis de engajamento do consumidor. Conteúdos com informações mais diretas, tais como preços e ofertas, estão associados a menores níveis de engajamento do consumidor. Por outro lado, uma combinação entre os dois tipos de conteúdos leva a maiores níveis de engajamento.
Lou e Yuan (2019)	Compreender o mecanismo pelo qual o <i>marketing</i> via ID afeta os consumidores.	O valor informativo do conteúdo criado por um ID, bem como sua credibilidade, atratividade e similaridade com seus seguidores estão associados a um maior nível de confiança por parte dos seguidores nos posts de publicidade de um ID.
Alalwan (2018)	Identificar e testar os principais fatores relacionados à publicidade via mídias sociais que possuem poder de predição das intenções de compras.	Os resultados indicaram que os fatores expectativa de desempenho, motivação hedônica, interatividade, informatividade e percepção de relevância, tiveram um efeito significativo na predição das intenções de compras.
Sokolova e Kefi (2020)	Investigar como a interação parassocial (IPS) entre audiência e ID das áreas de moda e beleza do <i>YouTube</i> e <i>Instagram</i> , em conjunto com a percepção de credibilidade destes, influenciam nas intenções de compras e como se relacionam com a atratividade social e física e o fator homofilia.	O fator homofilia está positivamente associado com a IPS, enquanto a atratividade física apresentou um comportamento contrário, estando negativamente associada à IPS ou não apresentando evidências de associação. Tanto a credibilidade do ID quanto a IPS se relacionaram positivamente com as intenções de compras.
Arora, Bansal, Kandpal, Aswani e Dwivedi (2019)	Propor um mecanismo para mensurar o índice de um ID nas mídias sociais mais populares, incluindo <i>Facebook</i> , <i>Twitter</i> e <i>Instagram</i> .	O engajamento, alcance, sentimento e crescimento têm um papel fundamental na determinação do índice de um ID.

Autor(es)	Objetivo	Principais resultados
Schouten, Janssen e Verspaget (2020)	Investigar a diferença entre o impacto da recomendação de uma marca/produto feita por uma celebridade tradicional e por um ID de mídia social.	De uma forma geral, os resultados mostraram que há uma maior eficácia de publicidade quando o endosso de um produto/marca é feito por um ID em comparação com uma celebridade tradicional.
Araujo, Neijens e Vliegenthart (2017)	Avaliar como diferentes tipos de usuários do <i>Twitter</i> podem influenciar a difusão de conteúdos de marcas famosas via <i>retweets</i> .	Os resultados indicaram que ID e corretores de informação estão associados a um maior número de <i>retweets</i> de conteúdos de marcas. A difusão desses conteúdos por parte de corretores de informação é amplificada quando os <i>tweets</i> originais fazem menção de ID.
Lin, Bruning e Swarna (2018)	Descrever como formadores de opinião <i>on-line</i> podem ser usados como ferramentas para a promoção de produtos e serviços.	O principal resultado é a proposição de um processo de planejamento em 5 estágios, podendo servir como um guia para empresários/marcas escolherem os endossadores ideais para seus produtos e serviços.
Hughes, Swaminathan e Brooks (2019)	Estudar os fatores que promovem o sucesso do engajamento <i>on-line</i> de uma marca com base no <i>marketing</i> via ID e <i>blogging</i> patrocinado.	Os resultados destacam o papel chave da interação entre os seguintes fatores: tipo de plataforma digital usada, intenção da campanha publicitária, fonte e tipo do conteúdo.

Fonte: Elaborado pelos autores (2020).

Kim e Kim (2018) examinaram o efeito do compartilhamento e da influência social no *Facebook* e *Twitter* nas vendas. Os autores usaram a regressão linear e descobriram que as referências sociais geradas pelo consumidor para os negócios aumentam significativamente as vendas no comércio social. No mesmo sentido, Tien, Rivas e Liao (2019) estudaram o impacto da influência de ID na decisão de compra e, por meio de equações estruturais, demonstraram que a persuasão e informação percebidas e o conhecimento da fonte são preditores adequados de influência dos influenciadores digitais.

Djafavora e Rushworth (2017) investigaram o impacto do *Instagram* na credibilidade da fonte, na intenção de compra do consumidor e na identificação social com diferentes tipos de celebridades. Os autores conduziram entrevistas em profundidade e os resultados da pesquisa mostraram que as celebridades no *Instagram* influenciam o comportamento de compra de jovens usuários, no entanto, as celebridades não tradicionais (influenciadores digitais) são mais poderosas, uma vez que os participantes da pesquisa percebem nos IDs uma maior credibilidade e identificação.

Hwang e Zhang (2018) usaram a modelagem de equações estruturais para explorar como as relações parassociais e a propaganda eletrônica boca a boca (eWOM) afetam a decisão de compra dos seguidores. Os resultados encontrados indicaram que o relacionamento parassocial modera os caminhos entre o conhecimento de persuasão dos seguidores e as intenções de compra e de eWOM. De maneira complementar, Sokolova e Kefi (2020) avaliaram como a credibilidade do blogueiro e a interação parassocial (PSI) estão relacionadas e influenciam a intenção de compra e, por meio de modelos de regressão, concluíram que a interação parassocial (PSI) exibe relações significativas e positivas com a intenção de compra.

Já Casaló, Flavián e Ibáñez-Sánchez (2018) identificaram alguns antecedentes e consequências principais da liderança de opinião na indústria da moda, usando modelos de regressão. Os autores concluíram que: aspectos como originalidade e singularidade influenciam positivamente a liderança da opinião no *Instagram*; a liderança de opinião afeta a intenção do consumidor de seguir os conselhos do influenciador; liderança de opinião afeta intenções comportamentais relacionadas ao influenciador; e, o ajuste percebido e a propensão à interação *on-line* atuam como moderadores.

Alalwan (2018), aplicando a modelagem de equações estruturais, concluiu que expectativa de desempenho, motivação hedônica, interatividade, informatividade e relevância percebida apresentam um impacto significativo na intenção de compra de uma cliente. Jiménez-Castillo e Sánchez-Fernández (2019), por sua vez, a partir de uma amostra de 280 seguidores, concluíram que poder influente percebido dos influenciadores digitais não apenas ajuda a gerar engajamento, mas também aumenta o valor esperado e a intenção comportamental em relação às marcas recomendadas. Além disso, o engajamento da marca no autoconceito aumenta o valor esperado da marca e as duas variáveis também afetam a intenção de comprar marcas recomendadas.

Ainda nesta temática, Dhanesh e Duthler (2019) examinaram os efeitos da conscientização de endossos pagos por influenciadores de mídia social no conhecimento de persuasão cognitiva dos seguidores (reconhecimento de anúncios), conhecimento de persuasão atitudinal (relacionamentos com o influenciador) e intenções comportamentais, especificamente a intenção da eWOM e intenção de compra. Os autores usaram modelos de correlação e regressão linear e concluíram que: a conscientização do endosso pago está correlacionada com o reconhecimento de anúncios e com o relacionamento influenciador-seguidor, que estão correlacionados com as intenções de compra e eWOM, no entanto, o reconhecimento de anúncios não afeta a relação influenciador-seguidor; a divulgação de

endosso pago melhora as percepções dos seguidores sobre a abertura e transparência do influenciador; as percepções de abertura e transparência do influenciador podem afetar positivamente as relações influenciador-seguidor e os resultados comportamentais

Arora *et al.* (2019) propuseram um mecanismo para medir o índice de influência dos IDs nas plataformas populares de mídia social (*Facebook*, *Twitter* e *Instagram*). Os resultados indicaram que o envolvimento, a divulgação, o sentimento e o crescimento desempenham um papel fundamental na determinação dos influenciadores.

3 Metodologia

3.1 Dados

Para atingir os objetivos deste estudo, foram consideradas séries temporais de vendas diárias de 10 empresas de luxo do setor do varejo de beleza e cosméticos da segunda maior loja de departamentos do Reino Unido. Os perfis oficiais das marcas no *Instagram* foram analisados e percebeu-se uma discrepância nos números apresentados, pois a marca com o maior número de seguidores na referida mídia social tem cerca de 30 milhões de usuários e a marca com menor número de seguidores apresenta cerca de 50 mil usuários. O porte e a trajetória comercial também são diversos, já que a amostra contempla marcas tradicionais e marcas que foram lançadas recentemente no mercado. Quanto aos influenciadores digitais, foram considerados neste estudo aqueles que, necessariamente, aparecem nas marcações do perfil oficial da marca. A maioria dos IDs gera conteúdos na área de beleza e cosméticos e apresenta um número de seguidores superior a 10 mil usuários. O fato de o número de seguidores não ser tão expressivo talvez se justifique por se tratar de marcas de luxo e, sendo assim, a popularidade dos influenciadores não se configura como requisito decisivo para a contratação. Assim, as informações das postagens nos perfis oficiais das marcas no *Instagram* e as publicações feitas por influenciadores digitais contratados que marcaram o perfil oficial da marca foram coletadas. O período analisado está compreendido entre os anos de 2017 e 2019.

O *Instagram* foi a mídia social escolhida, pois, segundo Highfield e Leaver (2014), é mais focado no aspecto comercial por meio da gestão eficiente de dados e da interação entre os usuários, o que possibilita uma publicidade com um alto nível de segmentação, incrementando o potencial gerador de *electronic word of mouth* (eWOM) e de interação com as marcas (De Veirman, Cauberghe, & Hudders, 2017). Ademais, diversos estudos ressaltam a importância da inclusão do *Instagram* em estudos acadêmicos, pois a rede possibilita a obtenção de informações relacionadas a questões sociais, culturais e ambientais sobre as atividades das

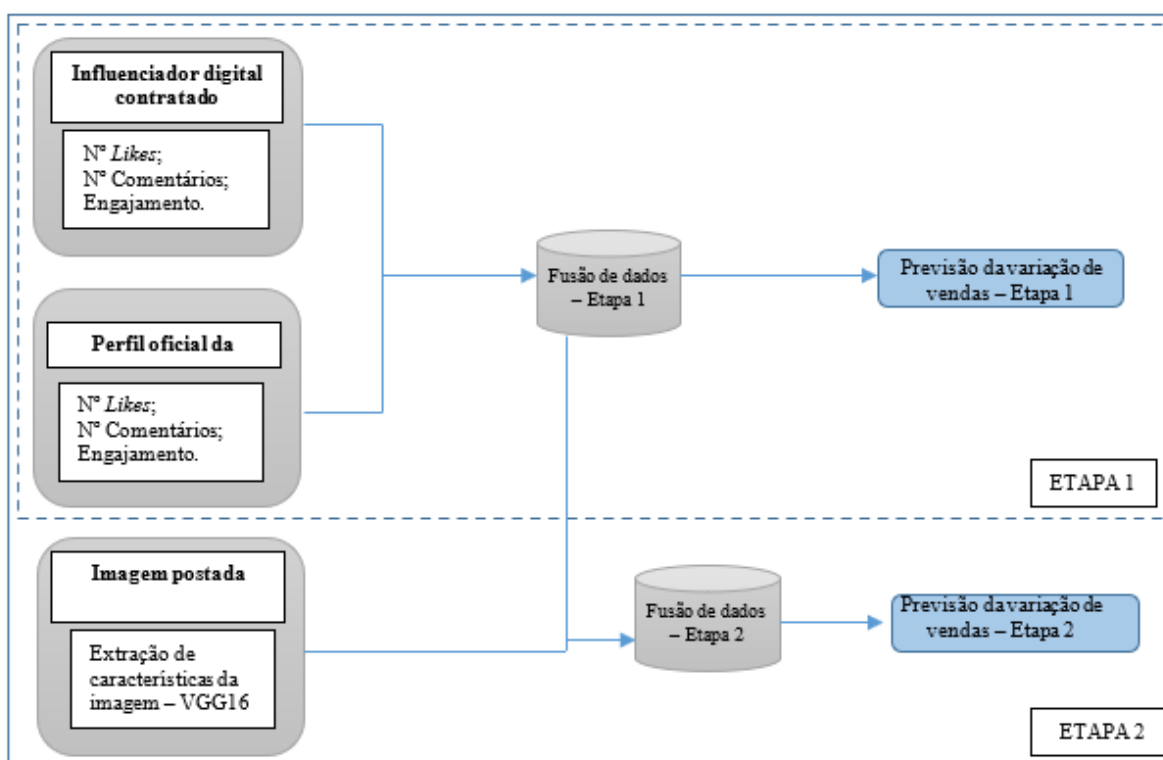
peças por meio de análises quantitativas e qualitativas sobre o comportamento dos usuários, (Hu, Manikonda, & Kambhampati, 2014; Lahuerta-Otero & Cordero-Gutiérrez, 2016).

Para avaliar a importância das mídias sociais, especificamente, das postagens feitas por influenciadores digitais no *Instagram*, os modelos de previsões univariados foram incrementados com variáveis relativas aos *posts*.

As variáveis relativas ao *Instagram* se subdividem em dois grupos específicos: o primeiro grupo de variáveis é composto pelos dados numéricos do perfil oficial da empresa estudada e da postagem feita pelo influenciador digital contratado (*likes*, comentários e engajamento); o segundo grupo é composto, basicamente, pela imagem postada pelo influenciador digital (Figura 1).

Figura 1

Modelo de Pesquisa – Etapas 1 e 2 de Previsão



Fonte: Elaborado pelos autores (2020).

Tendo em vista essa categorização de variáveis, a análise dos dados foi dividida em duas etapas, a saber (Figura 1):

- Etapa 1: desenvolvimento de modelos de previsão, considerando o comportamento histórico das variações diárias de receitas de vendas das dez empresas estudadas, com a inclusão das variáveis numéricas relativas às postagens dos influenciadores digitais e do perfil oficial da empresa (grupo de variáveis 1);

- Etapa 2: desenvolvimento de modelos de previsão a partir do comportamento histórico das variações diárias de receitas de vendas das dez empresas estudadas considerando a fusão de dados numéricos das postagens do Instagram e a extração de variáveis relacionadas às características das imagens postadas pelos influenciadores digitais (grupo de variáveis 1 e 2).

3.2 Métodos empregados

Compondo o núcleo da inteligência artificial, os modelos baseados em aprendizagem de máquina têm sido cada vez mais utilizados nos diversos campos do conhecimento. Na área de *marketing*, por exemplo, Lee *et al.* (2018) e Timoshenko e Hauser (2019) adotaram técnicas como o SVM (*Support Vector Machines*) e as Redes Neurais profundas para avaliar o engajamento e as necessidades do consumidor via mídias sociais, evidenciando a superioridade de desempenho dos modelos de aprendizagem de máquina sobre outros métodos tradicionais. Também lançando mão de dados provenientes de mídias sociais, Cui, Gallino, Moreno e Zhang (2018) demonstraram a utilidade dos modelos de aprendizagem de máquina na predição das vendas de uma empresa. Uma das razões para essa crescente importância e utilização de modelos de aprendizagem de máquina está justamente no poder que eles têm mostrado para aprender com grandes volumes de dados e identificar padrões (Jordan & Mitchell, 2015).

Em consonância com a literatura recente, a previsão das séries temporais de variações diárias de vendas das 10 empresas selecionadas neste trabalho foi feita por meio dos seguintes métodos, baseados em aprendizagem de máquina: *Support Vector Machines* (SVM) (Suykens & Vandewalle, 1999), *Random Forest* (Liaw & Wiener, 2002), XGBoost (Chen, He, Benesty, Khotilovich & Tang, 2015) e o método baseado em Redes Neurais profundas *Long Short-Term Memory*, LSTM (Simonyan & Zisserman, 2014).

O método SVM utiliza o conceito de planos de decisão para realizar as separações dos dados no espaço amostral, por meio de uma função *kernel* (Suykens & Vandewalle, 1999). Desse modo, o SVM cria grupos geométricos diferentes e reconhece padrões diferentes no espaço de dados. Após a definição geométrica do espaço de dados e o reconhecimento de padrões, é possível gerar inferências e realizar previsões (Suykens & Vandewalle, 1999).

Quanto mais diverso o conjunto de treinamento do modelo SVM, melhor serão os resultados de inferência.

O método *Random Forest* é composto por um conjunto de árvores de decisão (Liaw & Wiener, 2002). Nesse método, os dados são divididos em tabelas de decisões nas quais, conforme as decisões são tomadas, o algoritmo caminha entre os nós das árvores até que chegue à folha, então a inferência final é tomada. Esse método se destaca pela sua simplicidade e interpretabilidade, já que o usuário final consegue enxergar o porquê de cada decisão tomada pelo algoritmo (Liaw & Wiener, 2002).

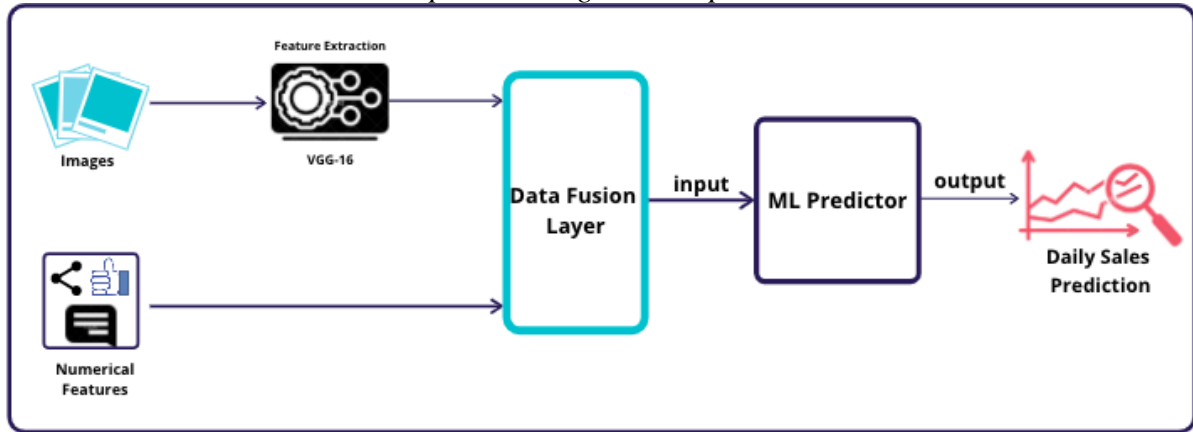
XGBoost é uma variação do algoritmo de árvores de decisão melhorado, por meio do método gradiente-descendente (Chen *et al.*, 2015), que tem uma alta taxa de acerto em dados estruturados devido a sua estrutura baseada *Gradient-boosting*.

LSTM é uma rede neural de aprendizagem profunda ou *Deep Learning* estruturada para a tarefa de previsão. É composta de uma estrutura recorrente, ou seja, em que a informação persista ao longo do tempo (componente temporal) (Simonyan & Zisserman, 2014). A rede LSTM é capaz de aprender dependências temporais, como também possui um fator de esquecimento que é ativado quando padrões não são utilizados durante um período. A rede LSTM se beneficiou do aumento da complexidade computacional para adição de um maior número de neurônios, possibilitando um aprendizado mais preciso (Simonyan & Zisserman, 2014). Além disso, a sua estrutura permite o trabalho paralelo com grandes conjuntos de dados, como os propostos nesta pesquisa.

Para a extração de características das imagens dos *posts*, foi utilizada a Rede Neural Convolutacional *VGG-16*. Segundo Simonyan e Zisserman (2014), a rede *VGG-16* se posiciona como estado da arte para a extração de característica de imagens. Neste trabalho, selecionaram-se os vetores de ativação (1X4096) da *fully connected layer* da rede *VGG-16* como entrada dos modelos, como ilustrado pela Figura 2.

Figura 2

Estrutura Geral do Modelo Proposto na Segunda Etapa da Análise



Fonte: Elaborado pelos autores (2020).

A camada de fusão de dados é responsável pela união dos dados numéricos com os dados provindos das imagens (*posts*), que são usados como entradas para alimentar os modelos de aprendizagem de máquina.

Para a avaliação dos modelos propostos foi utilizada a métrica *Root Mean Square Error* (RMSE) dada pela Equação (1).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (1)$$

Em que, n é o número de amostras, \hat{y} é o valor previsto. y é o valor real da série, e i é o índice de tempo da série temporal. A validação cruzada (*cross validation*) foi usada para medir o erro de precisão das previsões das séries temporais de variação de vendas diárias, a partir da seleção de pontos específicos de corte no histórico de dados. A RMSE é uma métrica interessante nesse contexto, pois mede a distância entre os valores previstos pelo modelo e os valores reais da série, penalizando grandes distâncias devido à diferença quadrática. Em outras palavras, quanto menor a RMSE, melhor o modelo.

Ademais, para testar se as diferenças estatísticas entre os modelos ajustados foram significativas, foi implementado um teste de análise de variância (ANOVA) com base nos RMSEs obtidos.

4 Análise dos resultados

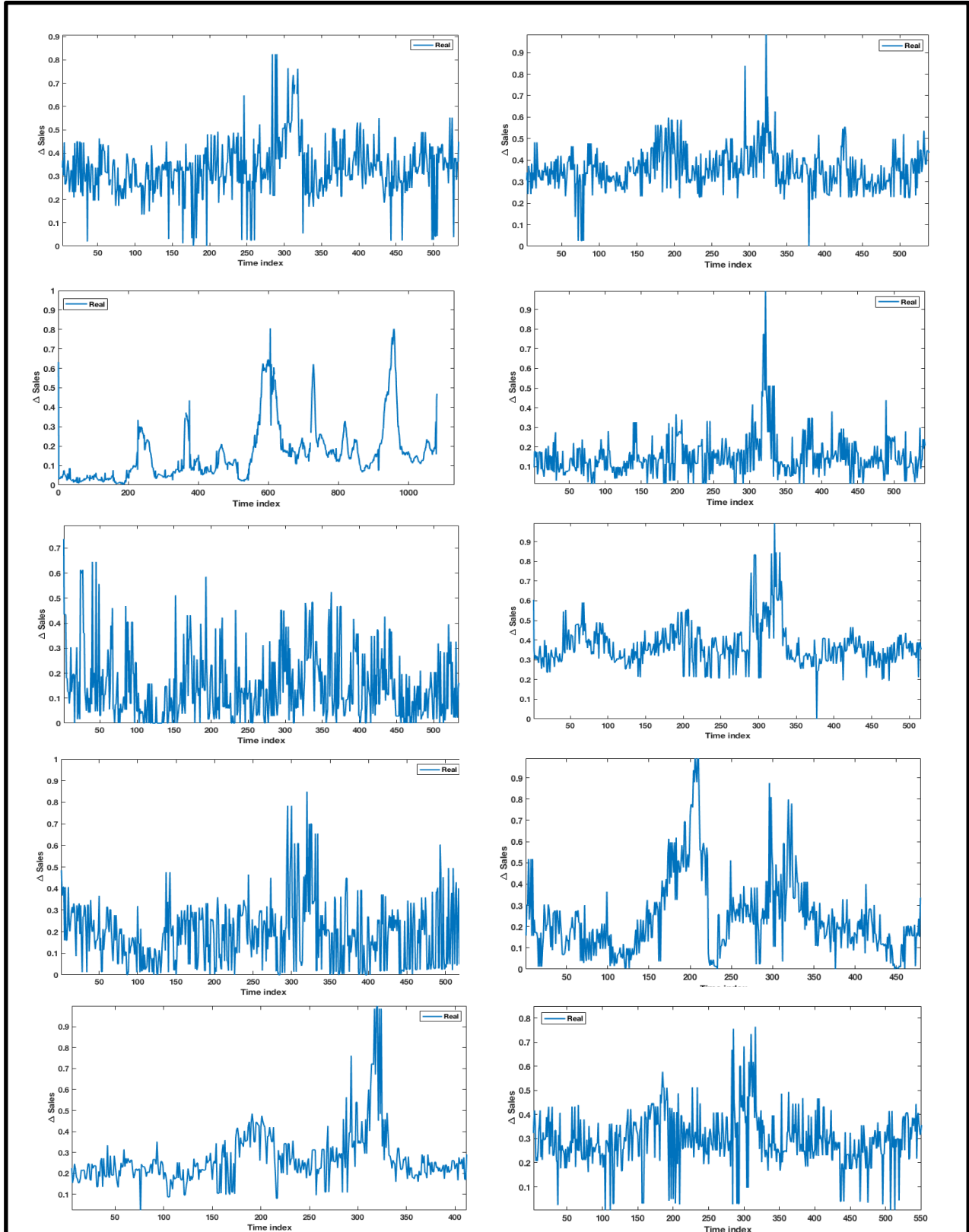
Para responder à questão norteadora deste trabalho e avaliar o poder preditivo das variáveis relacionadas às postagens feitas pelos influenciadores digitais (*Likes*, comentários, engajamento e imagens), apresentam-se, nesta seção, os resultados para as etapas de análises propostas neste estudo: i) previsão de variação diária de receita vendas a partir de variáveis numéricas provenientes de postagens do *Instagram*; ii) adição de imagens (*posts*) às previsões a partir da fusão de dados. Os modelos de previsão utilizados neste trabalho contemplam o estado da arte dos modelos de aprendizagem de máquina.

Inicialmente, são expostas as séries originais das variações da receita de vendas das 10 empresas estudadas (Figura 3). A variação da receita de vendas foi adotada para amenizar a variabilidade dos dados.

A partir da Figura 3, observa-se o comportamento histórico das variações diárias de vendas das empresas estudadas. A alta variabilidade dos dados pode ser explicada pelo caráter sazonal das estratégias de *marketing* adotadas pelas empresas para atrair e fidelizar os clientes. O lançamento de novos produtos e campanhas que priorizem a experiência do consumidor na loja física sintetizam as estratégias adotadas pelas empresas nos últimos anos.

Figura 3

Série da Variação das Receitas de Vendas das Empresas Estudadas



Fonte: Elaborado pelos autores (2020).

O caráter intervencionista das campanhas publicitárias, especialmente, por meio dos influenciadores digitais é demonstrado, quantitativamente, quando estes eventos se configuram como responsáveis pela quebra estrutural ou pela mudança de comportamento na série histórica de vendas. Neste estudo, os modelos de previsão de variações de vendas diárias com a inclusão de variáveis relativas às campanhas de *marketing* de influenciadores digitais serão capazes de avaliar o caráter intervencionista de tais estratégias.

4.1 Etapa 1 - Previsões da variação das vendas diárias considerando variáveis numéricas

Para cumprir com a primeira etapa de análise proposta neste trabalho, as séries de variações de vendas diárias foram utilizadas juntamente com as características numéricas dos *posts* dos influenciadores digitais e do perfil oficial da empresa (grupo de variáveis 1). Os resultados de previsões de variações diárias de vendas para as empresas consideradas no presente estudo são apresentados na Tabela 1.

Tabela 1

RMSE dos Modelos de Previsão Empregados

Método	Emp 1	Emp 2	Emp 3	Emp 4	Emp 5	Emp 6	Emp 7	Emp 8	Emp 9	Emp 10
LSTM	0.0504	0.0798	0.0517	0.1065	0.0898	0.0256	0.1374	0.0249	0.1137	0.0617
SVM	0.0641	0.1192	0.0582	0.1037	0.0906	0.1173	0.1247	0.0856	0.1201	0.0674
R. Forest	0.0689	0.1641	0.0631	0.1545	0.1238	0.0560	0.1735	0.0622	0.1756	0.0767
XGBoost	0.0945	0.1072	0.1329	0.1031	0.0998	0.1987	0.0994	0.1262	0.1436	0.1023

Fonte: Elaborado pelos autores (2020).

A partir da Tabela 1 é possível auferir que os modelos ajustados apresentaram um desempenho razoável na predição da variação de vendas das empresas analisadas. Entretanto, o método baseado em aprendizagem profunda, LSTM, demonstrou melhor resultado em 8 das 10 séries utilizadas. Sendo assim, os resultados demonstram a possibilidade de se prever vendas diárias a partir de dados provindos de mídias sociais. Neste caso, é possível afirmar que existe uma relação entre as estratégias empresariais no *Instagram*, especialmente, campanhas de

marketing por meio de influenciadores digitais e o desempenho em vendas das empresas analisadas.

Os resultados apresentados estão em concordância com os principais achados na literatura, evidenciando a superioridade de desempenho dos modelos de aprendizagem de máquina e os impactos do *marketing* via influenciadores digitais no comportamento do consumidor. Em especial, os resultados se alinham com o que foi apresentado por Cui *et al.* (2018). Os autores obtiveram um ganho de acurácia na predição das vendas diárias de uma empresa por meio do uso de modelos de aprendizagem de máquina em conjunto com dados provindos do *Facebook*. Ademais, o poder explicativo desses tipos de dados no comportamento do consumidor também foi corroborado por Lee *et al.* (2018) e Lou e Yuan (2019).

Ressalta-se, ainda, que os resultados encontrados estão coerentes com as evidências empíricas apresentadas por Kim e Kim (2018). Esses autores defendem que as referências sociais geradas pelos influenciadores digitais são importantes aspectos que afetam a decisão de compra por parte do consumidor. Nesse sentido, encontra-se coerência também com o estudo de Djafavora e Rushworth (2017), uma vez que os autores sinalizaram a importância de identificações sociais no processo de decisão de compra.

Na Figura 4, apresentam-se as séries temporais de variações de vendas diárias (linha azul) e as séries temporais previstas pelo método LSTM (linha vermelha). É possível notar, por meio das figuras, o grau de assertividade das previsões geradas pelo modelo de rede neural profunda.

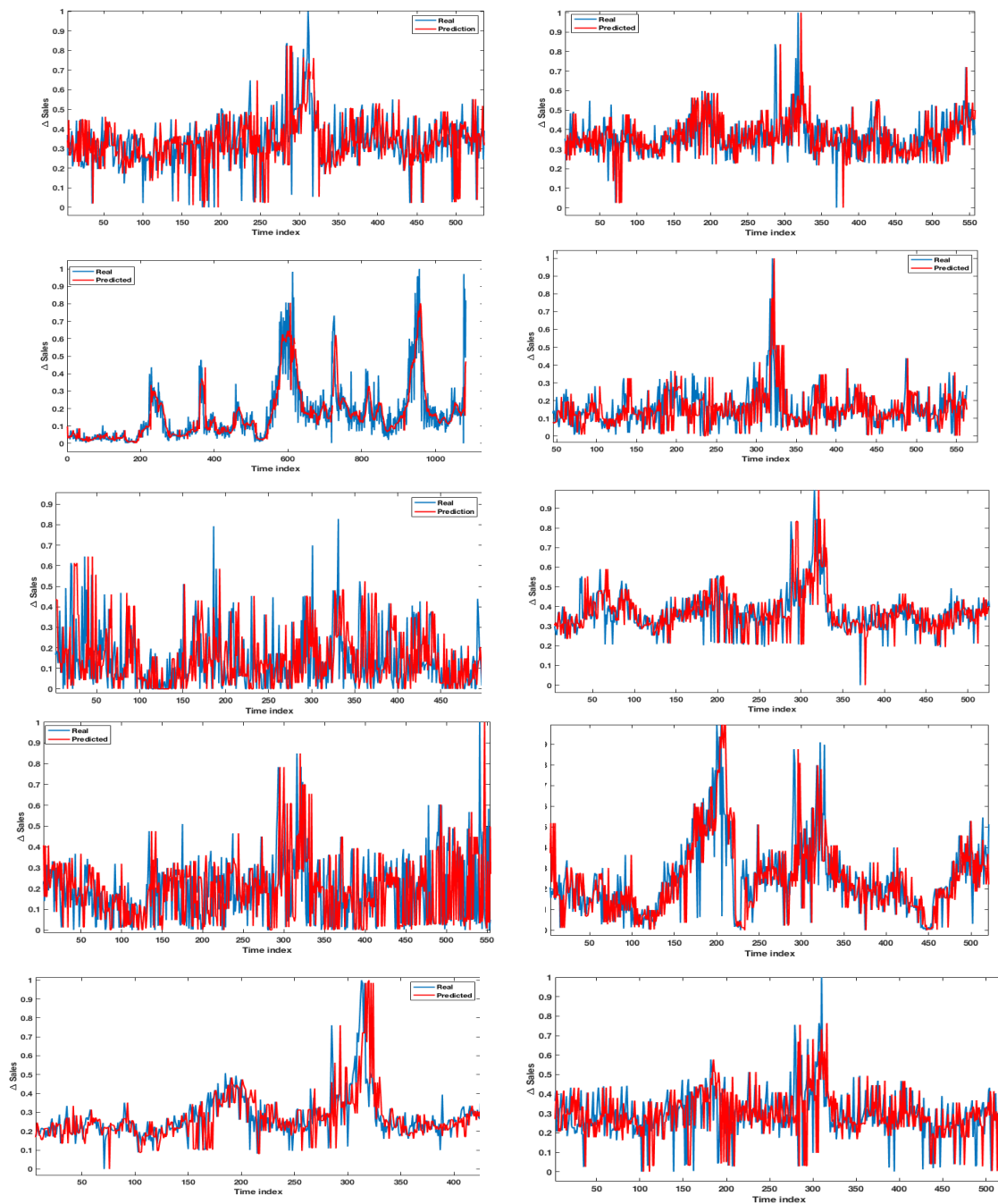
Além da possibilidade de utilizar dados do *Instagram* para a previsão de variações de vendas, os modelos estatísticos ajustados permitem a identificação das variáveis que mais contribuíram para o grau de acurácia das previsões geradas pelos modelos. Por meio da importância das variáveis fornecida pelo modelo XGBoost, é possível notar que o fator que mais influenciou nas previsões foram os *Likes* em *posts* em que a empresa foi marcada pelo influenciador digital. Esse fator se deu como mais importante para oito das dez empresas consideradas no estudo. Esta constatação reforça a ideia de que a identificação e referência social criadas pelo influenciador digital se configuram como uma eficiente estratégia de influência na decisão de compra do consumidor (Tien, Rivas, & Liao, 2019).

Para duas das empresas estudadas, a taxa de engajamento do influenciador digital se mostrou como a variável mais importante durante o processo de previsão. Esse resultado está em consonância com o estudo de Arora *et al.* (2019), no qual os autores mostraram que a variável engajamento foi uma das mais explicativas na determinação do influenciador digital.

Em linhas gerais, observa-se que as variáveis mais importantes no entendimento e previsão do padrão histórico de vendas dos casos analisados não estão relacionadas ao perfil oficial das empresas. As variáveis mais relevantes nos modelos de previsão pertencem ao perfil do influenciador digital contratado (*likes* e engajamento).

Figura 4

Previsão da Série Temporal Para as Empresas Consideradas no Estudo a Partir de Dados Numéricos Provindos das Redes Sociais



Fonte: Elaborado pelos autores (2020).

Sendo assim, a partir dos casos estudados, ressalta-se a relevância do uso de influenciadores digitais em campanhas de *marketing* e publicidade e a importância do conteúdo gerado por eles nas mídias sociais, uma vez que se demonstrou que os IDs podem gerar e/ou incrementar a conexão do público com a marca, mantendo um canal de comunicação aberto com os seguidores (Childers, Lemon, & Hoy, 2019).

4.2 Etapa 2 - Previsões da variação das vendas diárias considerando a fusão de dados

Na segunda etapa de análise, procedeu-se à fusão dos dados numéricos do *Instagram* (variáveis inseridas na primeira etapa de análise) com as imagens postadas pelos influenciadores digitais, conforme apresentado nas Figuras 1 e 2. O objetivo desta etapa de análise é verificar se as características extraídas das imagens postadas por meio da Rede Neural Convolutacional *VGG-16* contribuí para o entendimento do comportamento histórico das variações de vendas das empresas estudadas.

Em termos gerais, acredita-se que o reconhecimento dos padrões de características das imagens postadas tem um potencial de aprimoramento da acurácia dos modelos de previsão à medida que extrai e fornece ao algoritmo de aprendizagem informações exclusivas sobre as imagens e postagens que não podem ser facilmente encontrados em forma de dados. Para isso, os vetores de ativação (1X4096) da *fully connected layer* da rede *VGG-16* foram utilizados como entrada dos modelos juntamente com as variáveis numéricas. Os resultados desta etapa do estudo são apresentados na Tabela 2.

Tabela 2

RMSE dos Modelos de Previsão Empregados

Métodos de previsão	Emp 1	Emp 2	Emp 3	Emp 4	Emp 5	Emp 6	Emp 7	Emp 8	Emp 9	Emp 10
LSTM	0.0422	0.0615	0.0435	0.0898	0.0682	0.0512	0.0424	0.0382	0.0492	0.0430
SVM	0.0728	0.0985	0.0500	0.1002	0.0859	0.1145	0.2894	0.0631	0.0579	0.0954
R. Forest	0.0496	0.0635	0.0482	0.0976	0.0752	0.0745	0.1735	0.0622	0.1756	0.0767
XGBoost	0.0528	0.1285	0.0463	0.1421	0.0933	0.0721	0.0528	0.0441	0.0533	0.0521

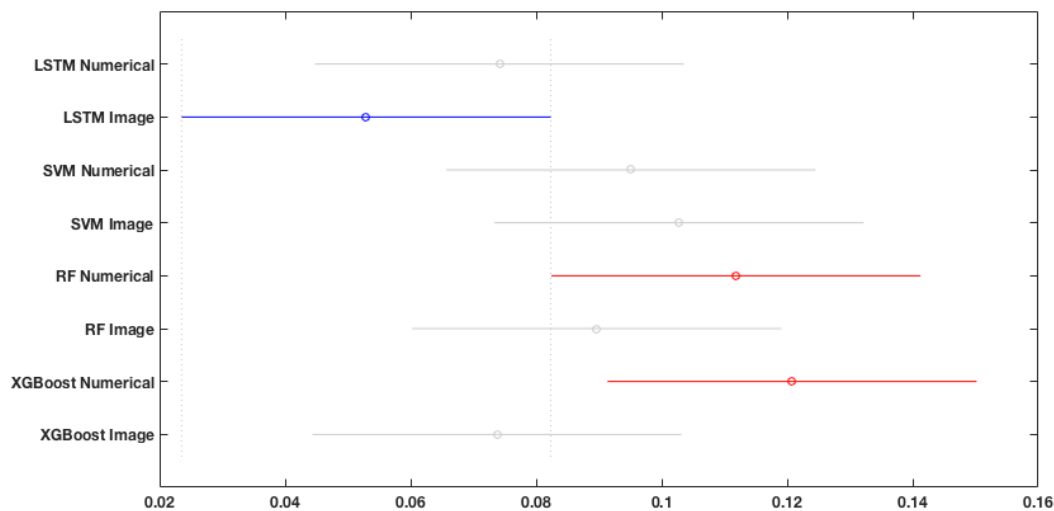
Fonte: Elaborado pelos autores (2020).

A partir das informações dispostas na Tabela 2, é possível notar que o método baseado em aprendizagem profunda (LSTM) obteve os melhores resultados em termos de acurácia de previsão quando comparados com os outros métodos tradicionais de aprendizagem de máquina. O resultado foi consensual para todas as empresas estudadas e tal fato pode ser explicado pela alta capacidade de aprendizagem do método, que é obtida por meio da utilização de milhões de neurônios, que são capazes de aproximar com maior eficiência os valores preditos dos valores reais de uma série temporal.

Ademais, é possível também notar que a extração e o acréscimo das características das imagens de cada *post* por meio da fusão das variáveis aprimoraram, de modo geral, as previsões elaboradas para as variações de vendas diárias das empresas. Demonstra-se, assim, que a imagem é um importante canal de comunicação que tem potencial impacto nas vendas das empresas estudadas. Esse resultado encontra respaldo nos achados empíricos de Bakhshi, Shamma e Gilbert (2014). Os autores demonstraram que a presença de imagens em mídias sociais incrementa o quantitativo de *likes* e comentários da postagem e engajamento do usuário e, conseqüentemente, aumentam a influência no processo de decisão de compra do consumidor.

Figura 5

Teste ANOVA Para Análise da Diferença Estatística Entre os Modelos Ajustados



Fonte: Elaborado pelos autores (2020).

O teste ANOVA, apresentado na Figura 5, foi realizado para todos os modelos de previsão ajustados tanto na etapa 1 (dados numéricos das postagens) quanto na etapa 2 (fusão de dados numéricos e imagens). Por meio dos resultados apresentados, evidenciou-se a diferença estatística dos modelos de previsão ajustados na etapa 2 em comparação com os modelos de previsão ajustados na etapa 1.

Por meio do teste ANOVA analisou-se, em termos gerais, a performance dos modelos de previsão ajustados e verificou-se a existência de diferenças estatísticas entre os diversos métodos e combinações de variáveis testados para todas as empresas. Sendo assim, esse resultado indica a eficiência em se utilizar imagens de *posts* como variável de entrada para se preverem variações de vendas diárias a partir de dados provindos de redes sociais, em específico o *Instagram*.

5 Considerações finais

Indubitavelmente, as mídias sociais se tornaram eficazes meios para a promoção de campanhas de *marketing* e para a conexão entre consumidores e marcas. Neste estudo, objetivou-se estudar a relevância das postagens no *Instagram* na construção de modelos de previsão de variação de receitas de vendas diárias. Para isso, as vendas diárias de dez empresas varejistas do setor de beleza e cosméticos foram analisadas.

Na primeira parte do estudo, o caráter preditivo das variáveis provenientes do *Instagram* foi testado. As séries de variações diárias de vendas das dez empresas foram modeladas considerando as características das postagens de influenciadores digitais e do perfil oficial da empresa. Os modelos empregados se mostraram eficientes na previsão e, por meio do XGBoost, foi possível observar que variáveis relativas ao perfil do influenciador digital são preponderantes para a qualidade da previsão. A importância das variáveis *likes* e engajamento reforça a ideia de que a identificação e referência social gerada pelo ID são importantes aspectos no processo de decisão de compra. Os resultados desta etapa estão coerentes com as evidências apresentadas por Cui et al. (2018), Lee et al. (2018) e Lou e Yuan (2019).

Na segunda etapa do estudo, foi proposta uma abordagem inovadora de fusão de dados numéricos com dados de imagem da postagem para o processo de previsão das séries de vendas. A partir do ajuste dos modelos, concluiu-se que as imagens são responsáveis por adicionar atributos exclusivos que ajudam na previsão e no entendimento dos padrões das séries de vendas analisadas. Pode-se afirmar que os modelos de *machine learning* incorporaram tanto as variáveis numéricas quanto apreenderam os padrões das imagens para gerar previsões mais

acuradas. Os resultados desta etapa da análise foram coerentes com os resultados empíricos encontrados por Bakhshi, Shamma e Gilbert (2014) e apresentaram uma extensão aos resultados encontrados pelos autores, ao utilizar e propor uma metodologia de extração e quantificação das características de uma imagem.

Em geral, o estudo demonstrou, de modo promissor, a eficiência dos métodos baseados em aprendizagem de máquina na previsão de variação de vendas diárias a partir de dados provindos do *Instagram*, especialmente, no que se refere à incorporação e extração de dados de imagens (análise visual quantitativa).

Ressalta-se a importância de estudos desta natureza para a formulação de estratégias de *marketing* mais eficientes, que permitam tanto a identificação de influenciadores digitais com maior potencial de influência e resultados quanto a formulação estratégica de postagens que tenham um maior impacto na audiência e na conversão em resultados de vendas por meio da utilização e extração de informações da análise de imagens. A combinação de tais fatores poderia gerar a maximização da influência e favorecer a ampliação do *marketing* viral por meio de redes sociais *on-line* de modo mais eficiente (More & Lingam, 2017).

As limitações deste trabalho estão relacionadas ao limitado número de empresas inseridas no estudo e a utilização de postagens feitas apenas no *Instagram* para o estudo do incremento do poder de previsão dos modelos estatísticos ajustados. Sendo assim, sugere-se, para trabalhos futuros, a utilização de outras mídias sociais como o *Twitter* e o *Facebook* para fins de comparação dos modelos de previsão. Além disso, o desenvolvimento de estudos desta natureza para outros setores seria de grande importância para a geração de novas evidências empíricas. Indica-se ainda a inclusão do conteúdo dos comentários das postagens, por meio da análise de sentimentos, como uma das variáveis de análise para os modelos de previsão. Acredita-se que a combinação de métodos possa enriquecer o processo e incrementar o grau de acurácia dos modelos desenvolvidos. Por fim, sugere-se a comparação entre o desempenho de influenciadores digitais e a publicidade profissional tradicional.

Referências

- Abolghasemi, M., Eshragh, A., Hurley, J., & Fahimnia, B. (2019). Demand Forecasting in the Presence of Systematic Events: Cases in Capturing Sales Promotions. arXiv preprint arXiv:1909.02716.
- Agnihotri, R., Dingus, R., Hu, M. Y., & Krush, M. T. (2016). Social media: Influencing customer satisfaction in B2B sales. *Industrial Marketing Management*, 53, 172-180. <https://doi.org/10.1016/j.indmarman.2015.09.003>
- Alalwan, A. A. (2018). Investigating the impact of social media advertising features on customer purchase intention. *International Journal of Information Management*, 42, 65-77. <https://doi.org/10.1016/j.ijinfomgt.2018.06.001>
- Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177-1190. <https://doi.org/10.1016/j.tele.2017.05.008>
- Araujo, T., Neijens, P., & Vliegenthart, R. (2017). Getting the word out on Twitter: The role of influencers, information brokers and strong ties in building word-of-mouth for brands. *International Journal of Advertising*, 36(3), 496-513. <https://doi.org/10.1080/02650487.2016.1173765>
- Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index-insights from Facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, 49, 86-101. <https://doi.org/10.1016/j.jretconser.2019.03.012>
- Asur, S., & Huberman, B.A. (2010). "Predicting the Future with Social Media", available at <http://www.hpl.hp.com/techreports/2010/HPL-2010-53.pdf>
- Audrezet, A., De Kerviler, G., & Moulard, J. G. (2018). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of Business Research*, 177, 557-569. <https://doi.org/10.1016/j.jbusres.2018.07.008>
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318. DOI: <https://doi.org/10.1509/jmr.14.0380>
- Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: photos with faces attract more likes and comments on Instagram. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '14*. Association for Computing Machinery, New York, NY, USA, 965–974. <https://doi.org/10.1145/2556288.2557403>

- Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2018). Influencers on Instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*, 177, 510-519. DOI: <https://doi.org/10.1016/j.jbusres.2018.07.005>
- Chen, I. F., & Lu, C. J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, 28(9), 2633-2647. DOI: <https://doi.org/10.1007/s00521-016-2215-x>
- Childers, C. C., Lemon, L. L., & Hoy, M. G. (2019). # Sponsored# Ad: Agency perspective on influencer marketing campaigns. *Journal of Current Issues & Research in Advertising*, 40(3), 258-274. DOI: <https://doi.org/10.1080/10641734.2018.1521113>
- Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1-4.
- Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management*, 27(10), 1749-1769. <https://doi.org/10.1111/poms.12707>
- De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. *International journal of advertising*, 36(5), 798-828. <https://doi.org/10.1080/02650487.2017.1348035>
- De Vries, N. J., & Carlson, J. (2014). Examining the drivers and brand performance implications of customer engagement with brands in the social media environment. *Journal of Brand Management*, 21(6), 495-515. <https://doi.org/10.1057/bm.2014.18>
- Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, 68, 1-7. <https://doi.org/10.1016/j.chb.2016.11.009>
- Dhanesh, G. S., & Duthler, G. (2019). Relationship management through social media influencers: Effects of followers' awareness of paid endorsement. *Public Relations Review*, 45(3), 101765. <https://doi.org/10.1016/j.pubrev.2019.03.002>
- Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*, 61, 47-55. <https://doi.org/10.1016/j.chb.2016.03.003>
- Felix, R., Rauschnabel, P. A., & Hinsch, C. (2017). Elements of strategic social media marketing: A holistic framework. *Journal of Business Research*, 70, 118-126. <https://doi.org/10.1016/j.jbusres.2016.05.001>
- Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570-576. <http://www.jstor.org/stable/20141547>

- Fildes, R., Goodwin, P., & Önkal, D. (2019). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, 35(1), 144-156. <https://doi.org/10.1016/j.ijforecast.2017.12.006>
- Fuchs, C. (2017). *Social media: A critical introduction*. (2nd ed.) London: Sage.
- Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing brands in the social media environment. *Journal of interactive marketing*, 27(4), 242-256. <https://doi.org/10.1016/j.intmar.2013.09.004>
- Gillon, K., Aral, S., Lin, C. Y., Mithas, S., & Zozulia, M. (2014). Business analytics: radical shift or incremental change? *Communications of the Association for Information Systems*, 34(1), 13. <https://doi.org/10.17705/1CAIS.03413>
- Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R., & Singh, R. (2016). Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. *Journal of business research*, 69(12), 5833-5841. <https://doi.org/10.1016/j.jbusres.2016.04.181>
- Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, 30(2), 127-135. [https://doi.org/10.1016/S0305-0483\(01\)00062-7](https://doi.org/10.1016/S0305-0483(01)00062-7)
- Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). The predictive power of online chatter. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 78-87). <https://doi.org/10.1145/1081870.1081883>
- Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, 56(3), 387-404. <https://doi.org/10.2501/IJMR-2014-025>
- Hajli, N., & Sims, J. (2015). Social commerce: The transfer of power from sellers to buyers. *Technological Forecasting and Social Change*, 94, 350-358. <https://doi.org/10.1016/j.techfore.2015.01.012>
- Highfield, T., Leaver, T. (2014). A methodology for mapping Instagram hashtags. *First Monday*, 20(1). DOI: <https://doi.org/10.5210/fm.v20i1.5563>
- Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What We Instagram: A First Analysis of Instagram Photo Content and User Types. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1). Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/14578>.
- Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2), 738-748. <https://doi.org/10.1016/j.ejor.2014.02.022>
- Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. (2016). The influence of social media interactions on consumer-brand relationships: A three-country study of brand perceptions

- and marketing behaviors. *International Journal of Research in Marketing*, 33(1), 27-41. <https://doi.org/10.1016/j.ijresmar.2015.06.004>
- Hughes, C., Swaminathan, V., & Brooks, G. (2019). Driving brand engagement through online social influencers: An empirical investigation of sponsored blogging campaigns. *Journal of Marketing*, 83(5), 78-96. <https://doi.org/10.1177/0022242919854374>
- Hwang, K., & Zhang, Q. (2018). Influence of parasocial relationship between digital celebrities and their followers on followers' purchase and electronic word-of-mouth intentions, and persuasion knowledge. *Computers in Human Behavior*, 87, 155-173. <https://doi.org/10.1016/j.chb.2018.05.029>
- Hyndman, R. J., & Athanasopoulos, G. (2014). Optimally reconciling forecasts in a hierarchy. *Foresight: The International Journal of Applied Forecasting*, (35), 42-48.
- Jiménez-Castillo, D., & Sánchez-Fernández, R. (2019). The role of digital influencers in brand recommendation: Examining their impact on engagement, expected value and purchase intention. *International Journal of Information Management*, 49, 366-376. <https://doi.org/10.1016/j.ijinfomgt.2019.07.009>
- Jin, S. V. (2018). "Celebrity 2.0 and beyond!" Effects of Facebook profile sources on social networking advertising. *Computers in Human Behavior*, 79, 154-168. DOI: <https://doi.org/10.1016/j.chb.2017.10.033>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- Kim, N., & Kim, W. (2018). Do your social media lead you to make social deal purchases? Consumer-generated social referrals for sales via social commerce. *International Journal of Information Management*, 39, 38-48. <https://doi.org/10.1016/j.ijinfomgt.2017.10.006>
- Kim, E., & Kim, Y.-C. (2018). Communication Infrastructure, Migrant Community Engagement, and Integrative Adaptation of Korean Chinese Migrants in Seoul. *Communication Research*, 1-23. <https://doi.org/10.1177/0093650218782819>
- Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*, 181, 145-153. DOI: 10.1016/j.ijpe.2015.09.011
- Kremer, M., Siemsen, E., & Thomas, D. J. (2016). The sum and its parts: Judgmental hierarchical forecasting. *Management Science*, 62(9), 2745-2764. <https://doi.org/10.1287/mnsc.2015.2259>
- Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing*, 80(1), 7-25. <https://doi.org/10.1509/jm.14.0249>
- Kulkarni, G., Kannan, P. K., & Moe, W. (2012). Using Online Search Data to Forecast New Product Sales. *Decision Support Systems*, 52 (3), 604-611. <https://doi.org/10.1016/j.dss.2011.10.017>

- Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, 64, 575-583. <https://doi.org/10.1016/j.chb.2016.07.035>
- Lamond, D., Dwyer, R., Ramanathan, R., Black, A., Nath, P., & Muyldermans, L. (2010). Impact of environmental regulations on innovation and performance in the UK industrial sector. *Management Decision*, 48(10), 1493-1513. <https://doi.org/10.1108/00251741011090298>
- Lassen, N. B., Madsen, R., & Vatrapu, R. (2014). Predicting iphone sales from iphone tweets. *2014 IEEE 18th International Enterprise Distributed Object Computing Conference* (pp. 81-90). IEEE. <https://doi.org/10.1109/EDOC.2014.20>
- Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of forecasting*, 22(3), 493-518. <https://doi.org/10.1016/j.ijforecast.2006.03.007>
- Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, 64(11), 5105-5131. <https://doi.org/10.1287/mnsc.2017.2902>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R news*, 2(3), 18-22. Retrieved from: <https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf>
- Lin, H. C., Bruning, P. F., & Swarna, H. (2018). Using online opinion leaders to promote the hedonic and utilitarian value of products and services. *Business Horizons*, 61(3), 431-442. <https://doi.org/10.1016/j.bushor.2018.01.010>
- Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. *International Journal of Information Management*, 35(4), 490-503. <https://doi.org/10.1016/j.ijinfomgt.2015.04.001>
- Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing*, 70(3), 74-89. <https://doi.org/10.1509/jmkg.70.3.074>
- Lou, C., & Yuan, S. (2019). Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, 19(1), 58-73. <https://doi.org/10.1080/15252019.2018.1533501>
- Lu, B., Fan, W., & Zhou, M. (2016). Social presence, trust, and social commerce purchase intention: An empirical research. *Computers in Human Behavior*, 56, 225-237. <https://doi.org/10.1016/j.chb.2015.11.057>
- More, J. S., & Lingam, C. (2019). A SI model for social media influencer maximization. *Applied Computing and Informatics*, 15(2), 102-108. <https://doi.org/10.1016/j.aci.2017.11.001>

- Oliva, R., & Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and operations Management*, 18(2), 138-151. <https://doi.org/10.1111/j.1937-5956.2009.01003.x>
- Ramanathan, U., & Muyltermans, L. (2011). Identifying the underlying structure of demand during promotions: A structural equation modelling approach. *Expert systems with applications*, 38(5), 5544-5552. <https://doi.org/10.1016/j.eswa.2010.10.082>
- Schivinski, B., & Dabrowski, D. (2016). The effect of social media communication on consumer perceptions of brands. *Journal of Marketing Communications*, 22(2), 189-214. <https://doi.org/10.1080/13527266.2013.871323>
- Schouten, A. P., Janssen, L., & Verspaget, M. (2020). Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit. *International journal of advertising*, 39(2), 258-281. <https://doi.org/10.1080/02650487.2019.1634898>
- Shiau, W. L., Dwivedi, Y. K., & Lai, H. H. (2018). Examining the core knowledge on facebook. *International Journal of Information Management*, 43, 52-63. <https://doi.org/10.1016/j.ijinfomgt.2018.06.006>
- Simonyan, K., & Zisserman, A. (2014). Very Deep convolutional networks for large-scale image recognition. *Computer Science*. arXiv preprint arXiv:1409.1556.
- Slack, N., Chambers, S., & Johnston, R. (2009). Administração da produção (Vol. 2). São Paulo: Atlas.
- Sokolova, K., & Kefi, H. (2020). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. *Journal of Retailing and Consumer Services*, 53, 101742. <https://doi.org/10.1016/j.jretconser.2019.01.011>
- Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300. DOI: 10.1023/A:1018628609742
- Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1), 1-26. <https://doi.org/10.1016/j.ejor.2015.11.010>
- Tien, D. H., Rivas, A. A. A., & Liao, Y. K. (2019). Examining the influence of customer-to-customer electronic word-of-mouth on purchase intention in social networking sites. *Asia Pacific Management Review*, 24(3), 238-249. <https://doi.org/10.1016/j.apmrv.2018.06.003>
- Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the operational Research Society*, 66(2), 299-307. <https://doi.org/10.1057/jors.2013.174>
- 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>