Received: 12 October 2019 | Revised: 24 August 2021 | Accepted: 12 March 2022

DOI: 10.1002/isaf.1508

RESEARCH ARTICLE

WILEY

Wikipedia pageviews as investors' attention indicator for Nasdaq

Raúl Gómez-Martínez | Carmen Orden-Cruz | Juan Gabriel Martínez-Navalón

Universidad Rey Juan Carlos, Madrid, Spain

Correspondence
Raúl Gómez-Martínez, Universidad Rey Juan Carlos, P.º de los Artilleros, 38, 28032 Madrid, Spain.
Email: raul.gomez.martinez@urjc.es

Summary

The attempt to measure investors' mood to find an early indicator of financial markets has evolved and developed with the advancement of technology over the years. The first attempts were based on surveys, a long and expensive process. Nowadays, big data has made it possible to measure the investor's mood accurately and almost entirely online. This paper analyzes the explanatory and predictive capacity of Wikipedia pageviews for the Nasdaq index. For this purpose, two econometric models have been developed. In both models, the explanatory variable is the number of Wikipedia visits, and the endogenous variable is Nasdaq index return. As an alternative to this approach, an algorithmic trading system has been developed. It uses Wikipedia visits as investment signals for long and short positions to check the predictability power of this indicator. It is determined that the volume of queries about Nasdaq companies is a statistically significant variable for expressing the evolution of this index. However, it has no predictive capacity. Keeping in mind the capacity of Wikipedia to exemplify Nasdaq trends, further studies should be conducted to determine how to make this indicator profitable.

KEYWORDS
algorithmic trading, behavioral finance, investors' mood, Wikipedia

1 | INTRODUCTION

For several decades, researchers have sought to measure investors' mood to anticipate market trends (Hilton, 2001). The first attempts were by Darling (1955), who used the relationship between dividends and profits to measure investors' mood. Another approach for measuring investors' moods is to use consumer confidence surveys (Lemmon & Portniaguina, 2006). Surveys have also been used to predict speculative bubble evolution (Shiller, 2000). However, surveys are expensive, and their results are only observable later. Therefore, new approaches must be considered. For example, Mnif et al. (2020) tried to measure investor sentiment using big data. They applied a search query of a list of words related to Islamic context and examined the engagement degree on social media and Twitter API (classified into positive and negative directions). Furthermore, Twitter has proved to be substantially useful in several areas. This is outlined by O'Leary (2015) in his analyses of the approaches that can be used to gather information and knowledge from Twitter.

Stock price prediction has been a constant area of financial research, using innovative techniques like machine learning (Nikou et al., 2019). Public opinion or mood has steadily gained importance over the past ten years due to the explosion of social media. Opinions have been converted into an instrument for business executives and academics because it is possible to quantify and qualify the "wisdom of crowds" (Pan et al., 2012). As of December 2018, there were 4.1 billion Internet users (Stevens, 2018), of which almost 3.5 billion were social media users (Datareportal.com, 2019). Both figures are equivalent to around half of the world's population. The US Census Bureau estimated the world population at 7.5 billion people in 2019 (Worldpopulationreview.com, 2019).

Individual user activity data has been converted into a database format, so that it can be utilized by various sectors. This data can be readily obtained from Google Trends, Twitter and Wikipedia. The search queries provide information about multiple topics. These databases have been used in medicine (Ginsberg et al., 2009) for early detection of disease activity. When followed by a rapid response, they can reduce the impact of both seasonal and pandemic influenza. However, these estimates do not have the precision that we would all like (Lazer et al., 2014). This suggests that disease incidence estimation models should incorporate not only data about how Internet features map to incidence but also additional data to estimate feature deceptiveness (Priedhorsky et al., 2019). In economics, engine data is used to forecast near-term values of economic indicators, such as automobile sales, unemployment claims, travel destination planning, consumer confidence (Choi & Varian, 2012), consumer behavior (Goel et al., 2010), and gasoline prices (Molnár & Bašta, 2017, June). In politics, it is used, for example, to predict elections results (Gayo-Avello et al., 2011; Shi et al., 2012). In education research, machine learning is used to investigate the role of open Internet knowledge through Wikipedia (Staub & Hodel, 2016).

In financial markets, these databases from Google Trends, Twitter, and Wikipedia have also been used to analyze trading behavior (Preis et al., 2013), prices (Challet & Ayed, 2013), returns (Joseph et al., 2011), stock market volatility (Dimpfl & Jank, 2016), liquidity, and volume (Bank et al., 2011). In addition, they have been used to predict future returns (Bijl et al., 2016; Kim et al., 2019) and investor sentiment. (Gómez-Martínez et al., 2019; Sprenger et al., 2014).

Moat et al. (2013) revealed that online data, extracted from Wikipedia, contained early information on stock market moves. Using edit logs for articles related to companies, they formed the Dow Jones Industrial Average (DJIA). Elshendy et al. (2018) noted that Wikipedia pages, along with the Global Data on Events, Language and Tone (GDELT) database and Twitter, have a high predictive power of crude oil prices. Cergol and Omladič (2015) found that Wikipedia page visits for individual companies have a great influence on future stock returns. They observed two different price patterns: a positive correlation, occurring solely in a bear market (denominated by them as a "merry frown") and a negative correlation only seen during a bull market (denominated as a "sour smile"); both phenomena explain the investors' corrective counterreaction to initial overpessimism/overoptimism. Kristoufek (2013) showed that the search queries and prices of bitcoin were connected and that high asymmetry exists between the effect of an increased interest in bitcoin while being above or below its trend value. Weng et al. (2018) revealed that a combination of online data sources (e.g., Google and Wikipedia) with traditional technical indicators provide a higher predictive power than any one source alone. They recognized that one of the limitations of their research was that they only examined one stock over a certain period (Apple).

This paper presents more evidence about Wikipedia as an explanatory and predictive variable of Nasdaq index performance. We tested the Wikipedia search frequency of several Nasdaq index stocks to see if they could be used to explain Nasdaq index performance. To accomplish this, a new model was constructed to forecast the stock market, using big data techniques based on this relationship. The model was tested using algorithmic trading strategies. This study methodology, which combines big data, artificial intelligence and algorithmic trading, covers three of the newest areas that are being used in the current research, but we rarely see them together in the same analysis.

Second 2 presents the hypothesis and the two methodologies used, Section 3 describes the data, Section 4 discusses the results, and Section 5 concludes.

2 | HYPOTHESIS AND METHODOLOGY

As described above, investors' mood can be measured following a big data approach using Tweet sentiments, Google Trends, or News analyzed employing natural language processing (NLP). Keeping this in mind, Weng et al. (2018) affirmed that Wikipedia has greater predictive power than Google. Therefore, in this paper, Wikipedia visits of listed companies' pages were used as an indicator of an investor's interest in those companies. The more visits received, the more investor interest could be assumed, which would then mean that an upward price trend should be observed.

Then, the hypothesis to test in this paper is:

Hypothesis 1. (H_1): Wikipedia visits of listed companies' pages cannot explain price evolution.

Hypothesis 2. (H_2): Wikipedia visits of listed companies' pages cannot predict price evolution for the next market season.

We propose three methodologies to validate these hypotheses:

i. ordinary least squares (OLS)
ii. algorithmic trading system (ATS) and
iii. artificial intelligence (AI) training a Bayesian network.

In the case of OLS, if the number of Wikipedia visits could explain or predict stock quotes, the β parameter of this model would be statistically significant. We define Model 1 with the following Equation 1

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \tag{1}$$

where:

- Y_t Monthly return of the Nasdaq 100 index (points monthly change in %)

- X_t Monthly percentage change of the total number of visits to the Wikipedia pages of the main companies listed in the index.

If the β parameter of Model 1 is positive and statistically significantly different from zero, H_1 is not validated.

To predict the index evolution, we propose Model 2, which is represented by Equation 2:

$$Y_t = \alpha + \beta X_{t-1} + \varepsilon_t, \tag{2}$$

where:

- Y_t Monthly return of the Nasdaq 100 index (points monthly change in %)
- X_{t-1} Monthly percentage change of the total number of visits to the Wikipedia pages of the main companies listed in the index in the prior period.

If the β parameter of Model 2 is positive and statistically significantly different from zero, H_2 is not validated.

In the case where β parameter of Model 1 or β parameter of Model 2 is statistically equal to zero, then H_1 and H_2 should be accepted, and Wikipedia visits would not have explanatory or predictability capacity.

However, it is possible that this predictor, although not statistically significant, is useful for obtaining profitability. Thus, we propose an alternative analysis methodology based on an algorithmic trading system using Wikipedia statistics. This system is bidirectional. Therefore, it takes long positions betting that the market will go up and short positions betting that the market will go down. We developed an algorithmic trading system that works in the following way:

1. If the monthly total of Wikipedia visits increases (or decreases), the system simulates a long (or short) position in the Nasdaq Futures market.
2. In the next month, we could find two possibilities:

a. If in the next month the number of visits increases, the system keeps the position open.
b. If in the next month the number of visits decreases, the system closes the position and opens a new position in the opposite way.

We can illustrate the operation of the system with an example:

- Suppose that the sum of visits to the Wikipedia pages of Nasdaq companies in February has been higher than that registered in January. This would be interpreted as an increase in the attention of investors on these companies, so it would be a bullish signal. Therefore, on February 1, the system opens a long position buying the future on the Nasdaq.
- Suppose that on February 28 we observe that the visits to the Wikipedia pages of the Nasdaq companies have been lower than in January. This is a bearish signal as investors' attention to these companies has diminished, so the system closes the open position and opens a new position, in this case short, selling the future on the Nasdaq.

And so on, as it is a swing system.

In the following section, we will validate H_2 if the system is not profitable and beats the market. Likewise, if the system is not profitable, we will accept hypothesis H_2.

Our third approach to measuring the predictive ability of Wikipedia is to train an artificial intelligence model. To do this, we developed a Bayesian network in which the predictors are the daily queries made in Wikipedia of the most representative companies of the Nasdaq by market capitalization. In contrast, the variable to predict will be what trend the Nasdaq will have in the next trading season, so it can only have two values (up or down).

In this case, we choose a daily frequency to have the maximum number of observations that allows us to train the model with the greatest precision.

We will reject H_2 retrospectively. We use 80% of the sample to train the model (the training dataset) and 20% is reserved as "clean data" to validate its predictive capacity (the validation dataset). If when testing the model with the "clean data" we achieve more than a 50% success rate for ups and downs, we will reject hypothesis H_2.

3 | DATA

We collected data from Wikipedia visits that were downloaded from the following webpage: https://tools.wmflabs.org/pageviews. Figure 1 shows how this tool plots the historical data of webpages visits. It is the largest online encyclopedia in the world, being multilingual, web based and free. The quality of information is verified by the research community (Staub & Hodel, 2016) and its reputation is covered by a standard procedure for complaints and its dispute-handling mechanism adheres to the principles of transparency (see contents in https://en.wikipedia.org/wiki/Help:Contents). This feature makes it a useful instrument for research.

The explanatory variable of the models (*X* variable) represents the change rate as a percentage of the sum of monthly visits to these Wikipedia pages, between one month and the last. These pages include Microsoft, Google, Apple, Amazon_(company), Facebook, Intel, Cisco_Systems, Comcast, Netflix, Adobe_Inc and Nasdaq (Nasdaq page and the Wikipedia pages of its the important companies listed).

Nasdaq index quotes were downloaded from the Investing webpage (https://www.investing.com/).

The dataset used for Models 1 and 2 regressions is shown in the Appendix.

The sample set started in January 2016 and ended in February 2019, resulting in a total of 38 observations. This period includes major events, such as the US Federal Reserve starting to increase interest rates, the Islamist terrorist attacks in Brussels (March 22, 2016), Brexit (June 23, 2016), the US presidential election

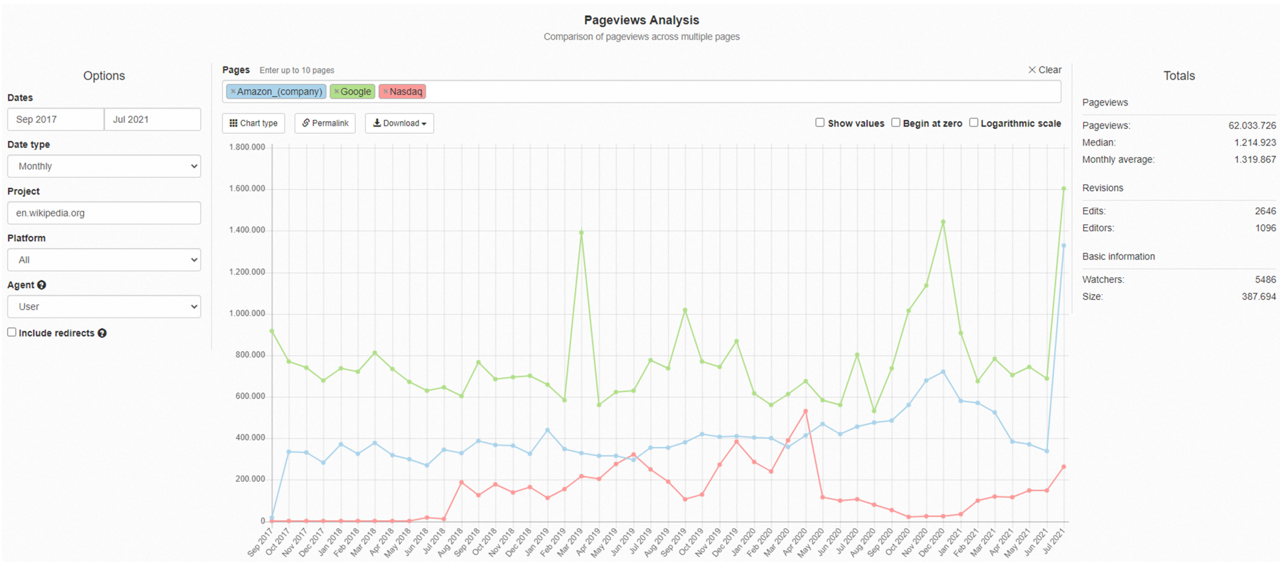

FIGURE 1 Wikipedia visits metrics *Source*: Wikimedia Foundation

TABLE 1 Model 1 OLS estimation

	Coefficient	Standard error	*t*-ratio	*p*-value
Constant	0.014	0.0065	2.12	0.0414 **
X	0.098	0.0436	2.26	0.0297 **
Statistics				
R^2	0.12		***F*(1,36)**	5.12
Adjusted R^2	0.10		***p*-value (*F*)**	0.03
Durbin–Watson	2.3			

Source: Authors' own research.

TABLE 2 Model 2 OLS estimation

	Coefficient	Standard error	*t*-ratio	*p*-value
Constant	0.013	0.0067	1.98	0.056 *
X	−0.051	0.0314	−1.64	0.1096
Stats				
R^2	0.07		***F*(1,36)**	2,69
Adjusted R^2	0.04		***p*-value (*F*)**	0.11
Durbin–Watson	2.4			

Source: Authors' own research.

(November 8, 2016), Islamist terrorist attacks in the USA and Spain (May 22 and August 17, 2017), North Korea nuclear threats (2017), summits between Trump and Kim Jong Un (2018), and a sizeable stock market decline (last quarter of 2018). Despite this, one of the most important events in the history of financial markets occurred in 2018, when Apple and Amazon hit 1 trillion USD in market capitalization (Sheetz, 2018). This most recent event highlights the importance of the technology sector. Apple and Amazon have been favorite investments for many investors over the past decade. This is due to their high returns, the way they have transformed their sectors, and the way they have continually acquired customers. The Nasdaq index was selected as the area of study because technology companies are a major component of the index and because these companies have become a major force in the economy.

4 | RESULTS

Models 1 and 2 parameters resulting from OLS methodology and by using the GRTL econometric tool are presented in Tables 1 and 2, respectively.

We can see that the β_1 parameter for Model 1 is positive and significant at the 95% confidence level. Therefore, Wikipedia visits can explain price movement in only 12% based on R^2 of the model. Therefore, we cannot validate H_1.

In contrast, parameter β_2 of Model 2 is not significantly different from zero. Therefore, we can validate H_2. In this case, Wikipedia visits have no predictive power.

In a disaggregated study, we can focus on the most important companies of the studied index that have been grouped under the acronym of FAAMG (Facebook, Apple, Amazon, Microsoft, and Google). Table 3 shows the estimated β parameters for Models 1 and 2. We see that none of them is significant, so we interpret that this disaggregated study is not relevant.

According to the alternative methodology proposed (ATS), different back tests were conducted. Back testing was performed using the Trading Motion SDK tool. Trading Motion is a fintech linked to iBroker and 24 other brokers across the world. It allows their clients to activate a portfolio from the 1,952 algorithmic trading systems, which have been developed by 85 different developers.[1] Each back test has a different rule to open a long position. For example, the first simulation shown in Table 4 opens a long position if Wikipedia visits vary from the previous month by more than 15%; otherwise, a short position is opened.

Table 3 shows that although there are some strategies with a positive return, none of them beat the market for the period from January 2016 to February 2019. Therefore, we can validate H_2

TABLE 3 Disaggregated *Model 1 and Model 2 β estimation*

	β_1	*t*-ratio	*p*-value	β_2	*t*-ratio	*p*-value
Facebook	0.00	0.71	0.48	0.00	0.51	0.61
Apple	0.00	0.29	0.77	0.00	−0.88	0.38
Amazon	0.00	0.15	0.88	0.00	−0.17	0.87
Microsoft	0.00	−0.40	0.69	0.00	1.37	0.18
Google	0.00	0.92	0.36	0.00	−0.05	0.96

Source: Authors' own research.

TABLE 4 Algorithmic trading simulations

	%	Net P&L ($)	Profit factor	Sharpe ratio	Annual ROI (%)	Success rate (%)	Capital suggested ($)	Capital required ($)
Long position for a percentage change higher than:	**15**	−66,477.00	0.83	−0.89	−6.55	28.21	315,000.00	10,450.00
	10	−23,162.00	0.94	−0.28	−3.06	35.90	235,000.00	10,450.00
	5	−18,795.00	0.95	−0.21	−2.65	38.46	220,000.00	10,450.00
	0	−57,922.00	0.85	−0.81	−6.66	35.90	270,000.00	10,450.00
	−5	16,195.00	1.05	0.28	3.24	43.59	155,000.00	10,450.00
	−10	19,290.00	1.06	0.39	7.04	53.85	85,000.00	10,450.00
	−15	52,455.00	1.16	0.88	12.52	64.10	130,000.00	10,450.00
	−20	47,989.00	1.15	0.79	11.46	64.10	130,000.00	10,450.00
	−25	47,989.00	1.15	0.79	11.46	64.10	130,000.00	10,450.00
	−30	55,803.00	1.17	0.92	13.32	68.42	130,000.00	10,450.00
	Market	57,858.00	1.18	0.96	13.81	65.79	130,000.00	10,450.00

Source: Authors' own research using Trading Motion SDK Tool.

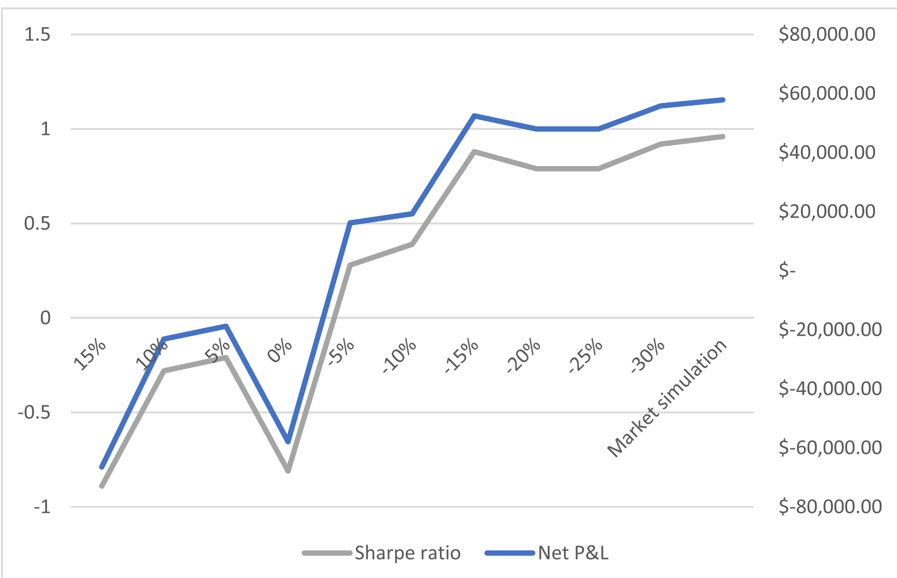

FIGURE 2 Performance of trading algorithmic systems based on Wikipedia

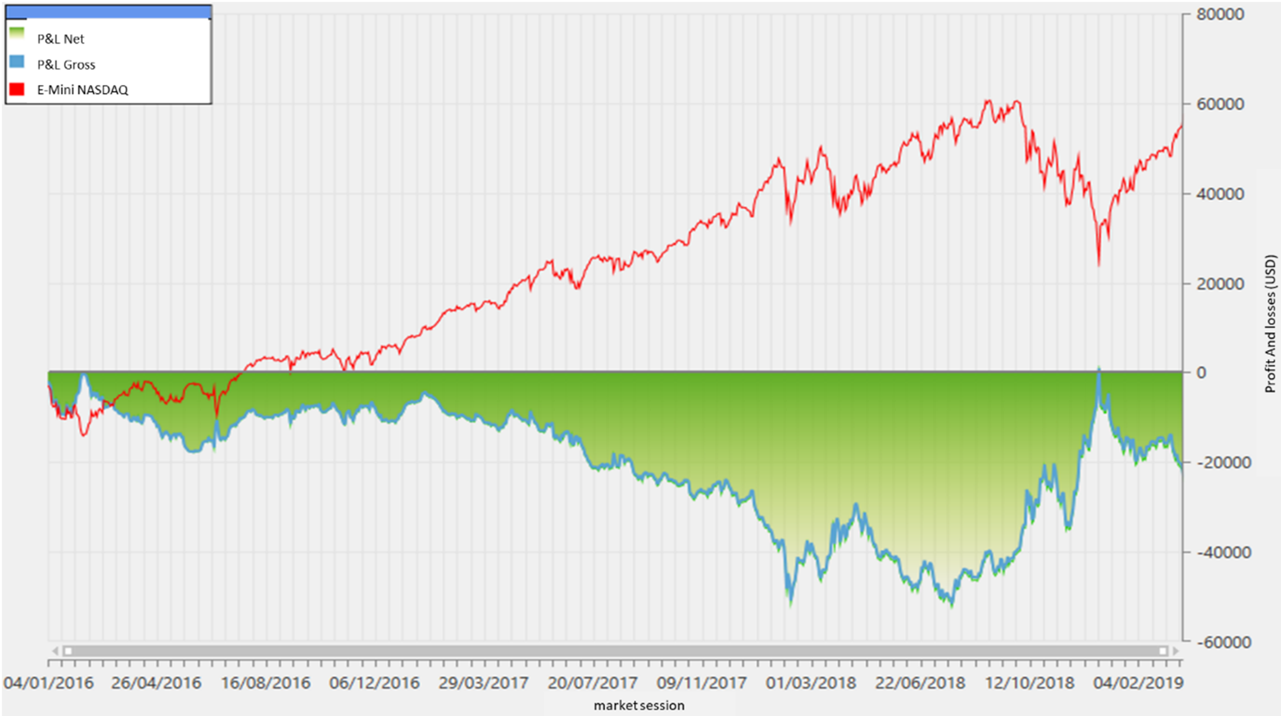

FIGURE 3 Profit and loss graph of the algorithmic trading system that open long position if Wikipedia visits variation is bigger than 0%
Source: Authors' own research using Trading Motion SDK Tool

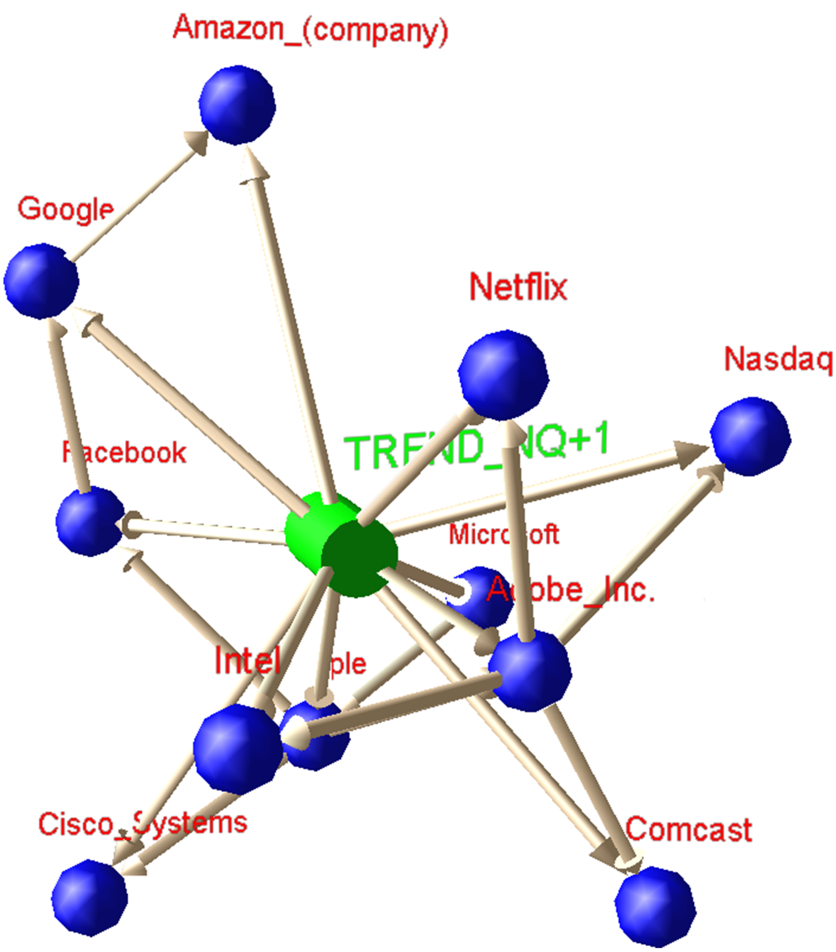

FIGURE 4 Bayesian Network trained based on Wikipedia

TABLE 5 *Bayesian Network validation*

	Success rate (%)	Prediction (%)	Test *F* (%)	ROC area (%)
Up	72.3	62.8	67.2	53.3
Down	34.2	44.6	38.8	53.3

Source: Authors' own research using dVelox.

TABLE 6 Bayesian Network confusion matrix

	Up prediction	Down prediction	Total
Up	81	31	112
Down	48	25	73
Total	129	56	185

Source: Authors' own research using dVelox.

using this alternative approach. Figure 2 shows the two main statistics of each trading system, the net profit and loss and the Sharpe ratio.

Figure 3 shows the profit and loss graph of the algorithmic trading system, which opened long positions if Wikipedia visits varied by more than 0%. This system is not profitable and does not beat the market (red line). Moreover, as shown in Figure 2, the systems that are closer to the market strategy, and open a few shorts position, are the ones resulting in a better performance.

The third methodology proposed in this study to validate the predictive capacity of visits to Wikipedia, a Bayesian network, does not show us enough predictive capacity of this indicator. The model has been trained using dVelox, an AI platform developed by Apara, an IT company.[2] The trained model is shown in Figure 4.

Table 5 shows the result of the validation of the model with 20% of the sample reserved, and Table 6 shows its confusion matrix.

Considering that the study period is mainly bullish, the model tends to overpredict the rises (up), observing a success rate far below 50% when the model predicts a fall (down). For this reason, we again accept hypothesis H_2, despite the model having an overall success rate of 57%.

5 | CONCLUSIONS

Predicting financial market performance is a heavily researched topic. Awareness of market trends is reflected on social media and can be measured through tracking indicators. In this paper, signals of awareness were collected from Wikipedia and translated in two models to explain and predict Nasdaq index performance. We found a positive correlation between both. Wikipedia visits, as a sentiment indicator about listed companies, define Nasdaq index evolution with a confidence of 95%. However, our models showed no predictive capacity, and the algorithmic trading systems do not outperform the Nasdaq index. Our research shows that Wikipedia visits, as a sentiment index about stocks, explain the evolution of the Nasdaq index, but they do not have enough predictive capacity in our models.

This study has several limitations related to the use of Wikipedia, as the variable studied was restricted to page view activity. We can count how many visits are made in Wikipedia but not if this visit is because we want to see if the company is going to go well or if the company is going to go badly. It is possible that other variables, such as the length and sentiment of each edit, the number of page edits or the number of users who contributed to a page could be useful explainers and predictors of Nasdaq index performance. This limitation will be addressed with further research.

ORCID

Raúl Gómez-Martínez https://orcid.org/0000-0003-3575-7970
Carmen Orden-Cruz https://orcid.org/0000-0002-1411-9286
Juan Gabriel Martínez-Navalón https://orcid.org/0000-0002-6836-6573

ENDNOTES

[1] For more information, visit https://www.tradingmotion.com/

[2] For more information visit: http://apara.es/en/

REFERENCES

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management*, *25*(3), 239–264. https://doi.org/10.1007/s11408-011-0165-y

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, *45*, 150–156. https://doi.org/10.1016/j.irfa.2016.03.015

Cergol, B., & Omladič, M. (2015). What can Wikipedia and Google tell us about stock prices under different market regimes? *ARS Mathematica Contemporanea*, *9*(2), 301–320. https://doi.org/10.26493/1855-3974.561.37f

Challet, D., & Ayed, A. B. H. (2013). Predicting financial markets with Google Trends and not so random keywords. arXiv preprint arXiv: 1307.4643. Retrieved from https://arxiv.org/pdf/1307.4643.pdf

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *The Economic Record*, *88*, 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Darling, P. G. (1955). A surrogative measure of business confidence and its relation to stock prices. *The Journal of Finance*, *10*(4), 442–458. https://doi.org/10.1111/j.1540-6261.1955.tb01297.x

Datareportal.com. (2019). Social media users by platform. Retrieved from https://datareportal.com/social-media-users

Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management*, *22*(2), 171–192. https://doi.org/10.1111/eufm.12058

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2018). Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, *44*(3), 408–421. https://doi.org/10.1177/0165551517698298

Gayo-Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). Limits of electoral predictions using twitter. In Fifth International AAAI Conference on Weblogs and Social Media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPaper/2862

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014. https://doi.org/10.1038/nature07634

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, *107*(41), 17486–17490. https://doi.org/10.1073/pnas.1005962107

Gómez-Martínez, R., Prado-Román, M., & Plaza-Casado, P. (2019). Big Data Algorithmic Trading Systems Based on Investors' Mood. *Journal of Behavioral Finance*, *20*(2), 227–238. https://doi.org/10.1080/15427560.2018.1506786

Hilton, D. J. (2001). The psychology of financial decision-making: Applications to trading, dealing, and investment analysis. *Journal of Psychology and Financial Markets*, *2*(1), 37–53. https://doi.org/10.1207/S15327760JPFM0201_4

Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, *27*(4), 1116–1127. https://doi.org/10.1016/j.ijforecast.2010.11.001

Kim, N., Lučivjanská, K., Molnár, P., & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*, *28*, 208–220. https://doi.org/10.1016/j.frl.2018.05.003

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, *3*, 3415. https://doi.org/10.1038/srep03415

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, *19*(4), 1499–1529. https://doi.org/10.1093/rfs/hhj038

Mnif, E., Jarboui, A., Hassan, M. K., & Mouakhar, K. (2020). Big data tools for ISLAMIC financial analysis. *Intelligent Systems Accounting Finance Management*, *27*(1), 1–12. https://doi.org/10.1002/isaf.1463

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific Reports*, *3*, 1801. https://doi.org/10.1038/srep01801

Molnár, P., & Bašta, M. (2017). Google searches and Gasoline prices. In 2017 14th *International Conference on the European Energy Market* (EEM) (pp. 1–5). IEEE.

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems Accounting Finance Management*, *26*(2), 1–11. https://doi.org/10.1002/isaf.1459

O'Leary, D. E. (2015). Twitter Mining for Discovery, Prediction and Causality: Applications and Methodologies. *Intelligent Systems Accounting Finance Management*, *22*, 227–247. https://doi.org/10.1002/isaf.1376

Pan, W., Altshuler, Y., & Pentland, A. (2012). Decoding social influence and the wisdom of the crowd in financial trading network. In 2012 *International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing* (pp. 203–209). IEEE.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, 1684. https://doi.org/10.1038/srep01684

Priedhorsky, R., Daughton, A. R., Barnard, M., O'Connell, F., & Osthus, D. (2019). Estimating influenza incidence using search query deceptiveness and generalized ridge regression. *PLoS Computational Biology*, *15*(10), e1007165. https://doi.org/10.1371/journal.pcbi.1007165

Sheetz, M. (November 20, 2018). Tech's 'FAANG' stocks have lost more than $1 trillion and counting from highs amid tech rout. CNBC, Markets. Retrieved from https://www.cnbc.com/2018/11/20/techs-popular-faang-stocks-have-lost-945-billion-and-counting-from-highs-amid-tech-rout.html

Shi, L., Agarwal, N., Agrawal, A., Garg, R., & Spoelstra, J. (2012). Predicting US primary elections with Twitter. https://snap.stanford.edu/social2012/papers/shi.pdf

Shiller, R. J. (2000). Measuring bubble expectations and investor confidence. *Journal of Psychology and Financial Markets*, *1*(1), 49–60. https://doi.org/10.1207/S15327760JPFM0101_05

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Staub, T., & Hodel, T. (2016). Wikipedia vs. academia: An investigation into the role of the Internet in education, with a special focus on Wikipedia. Universal. *Journal of Educational Research*, *4*(2), 349–354.

Stevens, J. (2018). Internet Stats & Facts for 2019. Hostingfacts.com site. Retrieved from https://hostingfacts.com/internet-facts-stats/

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, *112*, 258–273. https://doi.org/10.1016/j.eswa.2018.06.016

Worldpopulationreview.com. (2019). World Population by Country. Available in: http://worldpopulationreview.com/

How to cite this article: Gómez-Martínez, R., Orden-Cruz, C., & Martínez-Navalón, J. G. (2022). Wikipedia pageviews as investors' attention indicator for Nasdaq. *Intelligent Systems in Accounting, Finance and Management*, *29*(1), 41–49. https://doi.org/10.1002/isaf.1508

APPENDIX A

Dataset used for Model 1 and 2 regressions:

Date	Nasdaq	Wiki	%NQ	%WIKI
jul-15	4,588,91	2,839,197		
ago-15	4,274,58	2,826,040	−0.0685	−0.0046
sep-15	4,181,06	2,945,410	−0.0219	0.0422
oct-15	4,648,83	2,598,248	0.1119	−0.1179
nov-15	4,664,51	2,308,686	0.0034	−0.1114
dic-15	4,593,27	4,528,865	−0.0153	0.9617
ene-16	4,279,17	2,506,558	−0.0684	−0.4465
feb-16	4,201,12	2,422,423	−0.0182	−0.0336
mar-16	4,483,65	2,483,620	0.0673	0.0253
abr-16	4,341,3	2,159,573	−0.0317	−0.1305
may-16	4,523,89	2,096,911	0.0421	−0.029
jun-16	4,417,7	2,130,170	−0.0235	0.0159
jul-16	4,730,23	1,804,039	0.0707	−0.1531
ago-16	4,771,06	2,102,945	0.0086	0.1657
sep-16	4,875,7	2,287,561	0.0219	0.0878
oct-16	4,801,27	2,502,202	−0.0153	0.0938
nov-16	4,810,81	2,147,381	0.002	−0.1418
dic-16	4,863,62	2,325,864	0.011	0.0831
ene-17	5,116,77	2,237,528	0.052	−0.038
feb-17	5,330,31	2,088,520	0.0417	−0.0666
mar-17	5,436,23	2,399,120	0.0199	0.1487
abr-17	5,583,53	1,792,458	0.0271	−0.2529
may-17	5,788,8	2,172,816	0.0368	0.2122
jun-17	5,646,92	1,982,171	−0.0245	−0.0877
jul-17	5,880,33	1,915,659	0.0413	−0.0336
ago-17	5,988,6	2,260,028	0.0184	0.1798
sep-17	5,979,3	1,940,764	−0.0016	−0.1413
oct-17	6,248,56	2,180,754	0.045	0.1237
nov-17	6,365,56	2,043,326	0.0187	−0.063
dic-17	6,396,42	1,825,145	0.0048	−0.1068
ene-18	6,949,99	2,521,821	0.0865	0.3817
feb-18	6,854,42	2,128,869	−0.0138	−0.1558
mar-18	6,581,13	2,404,029	−0.0399	0.1293
abr-18	6,605,57	2,297,561	0.0037	−0.0443
may-18	6,967,73	2,158,227	0.0548	−0.0606
jun-18	7,040,8	1,980,385	0.0105	−0.0824
jul-18	7,231,98	2,185,600	0.0272	0.1036
ago-18	7,654,55	2,229,720	0.0584	0.0202
sep-18	7,627,65	2,061,629	−0.0035	−0.0754
oct-18	6,967,1	2,277,202	−0.0866	0.1046
nov-18	6,949,01	2,036,000	−0.0026	−0.1059
dic-18	6,329,97	1,757,982	-0.0891	-0.1366
ene-19	6,906,84	2,036,590	0.0911	0.1585
feb-19	7,117,01	1,686,558	0.0304	-0.1719