

Prediction analysis of the Herfindahl-Hirschman index based on recurrent neural networks

Cesar Hernandez^{*}, Fredy Martínez, Luis Pedraza

Universidad Distrital Francisco José de Caldas, Bogotá 111931, Colombia * **Corresponding author:** Cesar Hernandez, cahernandezs@udistrital.edu.co

CITATION

Article

Hernandez C, Martinez F, Pedraza L. (2024). Prediction analysis of the Herfindahl-Hirschman index based on recurrent neural networks. Journal of Infrastructure, Policy and Development. 8(13): 7795. https://doi.org/10.24294/jipd7795

ARTICLE INFO

Received: 4 July 2024 Accepted: 7 August 2024 Available online: 7 November 2024

COPYRIGHT



Copyright © 2024 by author(s). Journal of Infrastructure, Policy and Development is published by EnPress Publisher, LLC. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/ by/4.0/ **Abstract:** The technological development and growth of the telecommunications industry have had a great positive impact on the education, health, and economic sectors, among others. However, they have also increased rivalry between companies in the market to keep and acquire new customers. A lower level of market concentration is related to a higher level of competitiveness among companies in the sector that drives a country's socioeconomic development. To guarantee and improve the level of competition, it is necessary to monitor the concentration level in the telecommunications market to plan and develop appropriate strategies by governments. With this in mind, the present work aims to analyze the concentration prediction in the telecommunications market through recurrent neural networks and the Herfindahl-Hirschman index. The results show a slight gradual increase in competition in terms of traffic and access, while a more stable concentration level is observed in revenues.

Keywords: concentration index; HHI; traffic; revenue; access; telecommunications market; LSTM

1. Introduction

The services of the telecommunications market have a strong impact on important sectors of society such as education with access to the internet and information and communication technologies, health with the strengthening of telemedicine, and the economy through technologies that increase productivity, among others. The installation and deployment of fifth-generation mobile technologies (5G); as support for next-generation services such as the Internet of Things (IoT) and even talk of Industry 4.0 as a leading trend of the "fourth industrial revolution," are based on telecommunications services provided through ICT (Bardey et. al., 2022; CRC, 2023; OECD, 2023; Ilchenko et. al., 2019).

The development of the telecommunications industry has been experiencing strong growth driven by technological innovation and strengthened by market demand and government policies (Mehrotra and Menon, 2021; Lin et. al., 2018). The current demand for the telecommunications market has increased the necessary bandwidth due to the new services and applications that have been developed (Canadian Radio-television and Telecommunications Commission, 2023). This increases the sales volume and revenues of telecommunications companies and Internet Service Providers (ISPs), producing a strong rivalry between companies to keep and win new customers (Fiedler et. al., 2017).

The preceding highlights the importance and need to conduct studies and analysis of competition in the telecommunications market, aiming to monitor and control market concentration to avoid the emergence of monopolies and oligopolies (Apolinario et al., 2023). This makes it possible to improve competition, reduce the digital divide and strengthen the development of the countries' economies. A lower level of market concentration is related to a higher level of competitiveness that drives socioeconomic development, which brings with it a greater quantity and quality of benefits for society and has a positive and large impact on areas such as education, health, and the economy (Commission for the Regulation of Communications, 2022; Melnik, 2008; Miera Berglind, 2015).

All of the above highlights the importance of monitoring the concentration level of the telecommunications market. However, predicting its behavior would be even better, since this would allow planning in advance the necessary strategies to regulate this market (Amin et. al., 2018; MinTIC, 2024; Nekmahmud, 2018). This paper aims to analyze concentration prediction in the telecommunications market. To achieve this, we worked with real data from companies in the Colombian telecommunications market; between 2012 and 2022, the Herfindahl-Hirschman index (HHI) was used to measure the level of concentration of companies in the telecommunications sector, and the prediction of this index was made through Long Short-Term Memory (LSTM) recurrent neural networks (Berradi et al., 2021; OFCOM 2022).

This work is organized and presented in five sections. Section 2 presents the research methodology. Section 3 describes the results obtained. Section 4 discusses the results. The final section, section 5, establishes the general conclusions of the work.

2. Materials and methods

The concentration analysis was initially carried out through the Herfindahl-Hirschman index (HHI) to predict concentration in the telecommunications market. The HHI was calculated for each year using metrics (traffic, revenue, access). Subsequently, the predictive model based on recurrent neural networks Long Short-Term Memory (LSTM) was developed:

- Since the data are time series, a suitable option would be to use specific models for this data type, such as ARIMA or Prophet. However, recurrent Long-Short-Term Memory (LSTM) neural networks are considered here to identify more complex patterns in the data;
- Once the model is trained, predictions are made and visualized to see how the market is expected to behave.

2.1. Data collection process (traffic, revenue and accesses)

The experimental data were obtained through the Communications Regulation Commission (CRC), which is the regulatory authority for Colombia. The CRC provides access to relevant data from telecommunications companies operating in the Colombian market over the last decade through its website (https://www.postdata.gov.co/).

For this work, it was decided to select the variables of traffic, income and access, with which an adequate and fair analysis of the concentration of the Colombian telecommunications market can be carried out. The traffic, income and access data are taken monthly for the period between 2012 and 2022, for each of the telecommunications companies in force during said period. Another point to highlight

is that, although the CRC provides the previous information for the prepaid or postpaid modality, in this document it was decided to work with the global conglomerate of the data, that is, the sum of prepaid and postpaid.

The experimental design begins by extracting the data information on traffic, income, and access variables from the official CRC page. Subsequently, the data is cleaned and preprocessed. Then, the HHI index is built. Then, the necessary parameters for the LSTM prediction model are calculated. Once the LSTM model has been developed, forecasts are made for the following periods. The error level is determined and the corresponding adjustments are made if necessary. The corresponding graphs are built, and the results are analyzed.

2.2. Herfindahl-Hirschman Index (HHI) for each year and metric (traffic, revenue, accesses)

The Herfindahl-Hirschman Index (HHI) is calculated by adding the squares of the market shares of all companies in the market, see Equation (1) (Communications Regulatory Commission, 2022; Lis-Gutiérrez, 2023).

$$HHI = \sum_{i=1}^{n} s_i^2 \tag{1}$$

where if is the market share of the firm i and n is the total number of firms in the market. We'll calculate the HHI for each metric (traffic, revenue, access) and year.

The traffic, revenue, and access variables data for each Colombian telecommunications sector company were obtained from (Communications Regulatory Commission, 2023). **Table 1** describes the HHI values for the traffic variable, while the HHI values for the revenue and access variables can be seen in **Tables 2** and **3**, respectively. These values can also be seen graphed in **Figure 1**.

Year	Average HHI	HHI Maximum	HHI Minimum
2012	3515.39	5053.23	3262.86
2013	3099.38	3256.12	2961.36
2014	3068.55	3320.02	2544.32
2015	3326.15	3686.38	3157.35
2016	3573.07	3733.86	3396.86
2017	3487.93	3602.19	3379.42
2018	3548.15	3665.56	3404.11
2019	3259.34	3383.05	3094.41
2020	3569.88	3743.21	3208.86
2021	3197.75	3554.63	3026.10
2022	2925.52	3041.91	2873.54

Table 1. HHI Index values for traffic.

Year	Average HHI	HHI Maximum	HHI Minimum		
2012	3737.51	5045.73	3361.15		
2013	3521.80	3583.97	3422.97		
2014	3566.59	3626.51	3507.79		
2015	3881.58	3954.15	3694.53		
2016	3925.36	4020.40	3838.79		
2017	3947.45	4056.83	3794.11		
2018	3800.42	3913.56	3727.18		
2019	4130.75	4262.81	3907.81		
2020	4432.90	4547.29	4321.43		
2021	4428.48	4559.45	4303.73		
2022	4251.42	4389.50	4162.31		

Table 2. HHI Index values for revenue.

Table 3. HHI Index values for the accesses variable.

Year	Average HHI	HHI Maximum	HHI Minimum
2012	4251.28	5621.05	3572.08
2013	4401.70	4618.33	4204.62
2014	4959.30	5366.19	4515.88
2015	4236.18	4687.51	4012.11
2016	4198.22	4359.88	4053.61
2017	4028.05	4288.64	3911.63
2018	3772.31	3935.02	3619.29
2019	3731.01	3830.15	3604.40
2020	3846.43	3943.86	3763.44
2021	3742.88	3897.52	3614.55
2022	3598.99	3627.33	3561.65



Figure 1. HHI index over time.

2.3. Predictive LSTM model

A multi-input Long Short-Term Memory (LSTM) neural network approach is proposed, optimized for time series tasks to analyze and predict the behavior of the telecommunications market in Colombia. A model will be designed to individually process three key metrics: traffic, revenue, and access. Each metric will be channeled through its own LSTM layer, and subsequently, the outputs from these layers will be combined for a unified interpretation. This multi-input architecture captures the interdependencies and individual patterns in each metric. Moreover, data from all service providers will be aggregated to project the overall market trend. This approach aims to attain more accurate and comprehensive predictions of the market's future behavior, based on historical trends.

Building upon the proposed multi-input LSTM neural network approach, the preprocessing pipeline begins with data normalization. This ensures that all data values fall within the range [0, 1], a step deemed essential to promote the efficient training of neural networks, and to avoid undue dominance of any particular feature during the learning phase.

Sequential data structure is paramount for LSTMs. To this end, sequences are formed using a sliding window approach, where a 3-year window serves as the input to predict the subsequent year. This method captures the immediate past and considers longer-term trends, enhancing the model's predictive power.

The choice of 50 units in the LSTM layer is a strategic balance. While increasing the number of units could potentially enhance the network's capacity to model intricate patterns, it could also risk overfitting, especially when data is limited. The 'relu' activation function in the LSTM layer stands out due to its capability to accelerate convergence and reduce the vanishing gradient issue, a common challenge in deep networks.

Following individual processing, the outputs from the LSTM layers are amalgamated. This combined architecture ensures the model benefits from individual metric patterns and interdependencies. The subsequent dense layer outputs the prediction. The Adam optimizer, recognized for its robustness and efficiency in regression tasks, is chosen with a learning rate of 0.01. This learning rate, albeit higher than conventional defaults, expedites convergence without destabilizing the learning process. Training is conducted over 200 epochs, determined experimentally to ensure a comprehensive understanding of the underlying patterns while preventing overfitting. (see Figures 2–7).

Interventional studies involving animals or humans, as well as other studies that require ethical approval, must list the authority that provided approval and the corresponding ethical approval code.



Figure 4. Loss for accesses.



A joint normalization approach was adopted to visualize the three-time series (traffic, revenues, and accesses) coherently. All series were scaled to the range [0, 1] using a single scaler, based on the global range of the combined data. This strategy

ensures that the relative differences between series are retained, allowing for effective visual comparison. Once normalized, the previously trained LSTM models were used to project the three metrics over the next five years. These projections were based on the most recent data from each series. Finally, the predictions were visualized in a single plot, highlighting the projections for the next five years for traffic, revenues, and accesses (See **Figure 8**). This representation provides a clear insight into the anticipated trends for the telecommunications market in Colombia.



The five-year projections using LSTM models for traffic, revenues, and access metrics reveal distinct trends for each time series. Firstly, the traffic metric exhibits an upward trend, potentially indicating a greater usage of telecommunication resources in the future. In contrast, revenues seem to decrease slightly, which might reflect heightened competition in the sector or the introduction of more affordable solutions for consumers. Lastly, the number of accesses appears to plateau, suggesting a possible saturation point in the market or a consolidation of service providers. These projections, grounded on historical patterns, underscore the importance of continuously monitoring the market and adjusting business strategies in line with anticipated trends.

2.4. LSTM multi-input forecasting with traffic-revenue-access integration

In striving for a robust and accurate traffic prediction, it's imperative to consider multiple factors that might influence this metric. Thus, we've adopted an approach that incorporates not only traffic history but also the revenues and access of each company. This choice is grounded in the premise that past and present behavior in these three areas can offer critical insights into future trends. The multi-input LSTM network architecture has been specifically chosen for its inherent ability to recognize and learn long-term patterns in time series. Moreover, by segregating the time series into individual LSTM layers, the model can learn unique patterns within each series and subsequently merge this information in the dense layers for a cohesive interpretation. This composition lets the network grasp interdependencies and mutual influences between traffic, revenues, and access. Regarding the training strategy, choosing a 'relu' activation function in the dense layers is based on its proven efficacy in mitigating the vanishing gradient problem. At the same time, the Adam optimizer is employed for its capability to dynamically adjust the learning rate, facilitating a quicker and steadier convergence. See **Figure 9**.



Figure 9. The architecture of the multi-input LSTM neural network for predicting traffic based on traffic, revenue, and access data.

3. Results

In **Figures 10–24**, you can see the behavior of the HHI index for the traffic of each company in the Colombian telecommunications market from 2012 to 2022, together with the behavior of the prediction made by the proposed LSTM model. It is important to mention that although there are 14 telecommunications companies, not all have operated throughout the eleven years of observation. Only three have remained; the others have entered later and endured or entered and left.



Figure 10. HHI Index Prediction for Enterprise Traffic 1.







Figure 14. HHI Index Prediction for Enterprise Traffic 4.























Figure 23. HHI Index Prediction for Enterprise Traffic 13.



Figure 24. HHI Index Prediction for Enterprise Traffic 14.

4. Discussion

This section discusses the analysis of the results obtained. One of the most important limitations of this work was the inconsistency of the participation of several telecommunications companies during the period analyzed. Likewise, the time range of information, corresponding to a decade with monthly data, could be greater to have a broader overview of the telecommunications market. Finally, the data corresponds to the two most recent years. This is because the amount of data available to train the model is key for machine learning models.

4.1. Analysis of HHI behavior

Over the past decade, the HHI in the Colombian telecommunications market has shown varied concentration across traffic, revenue, and access metrics. All metrics have remained above the 2500 threshold, indicating a highly concentrated market.

- Traffic: It started with an HHI of 3515 in 2012 and, despite some fluctuations, has decreased to 2926 in 2022. This suggests a gradual decline in concentration in terms of traffic over the decade, though it remains a highly concentrated market.
- Revenue: It began with an HHI of 3738 in 2012 and saw an increase in concentration by 2020 with an HHI of 4433, only to decrease in 2022 to a value of 4251.
- Access: This metric has had the highest and most fluctuating HHI, starting at 4251 in 2012, peaking at 4959 in 2014, and decreasing to 3599 in 2022.

4.2. Determination of correlations between curves

I calculate the Pearson correlation coefficient to determine the correlation between the HHI curves of traffic, revenue, and access. This coefficient varies between -1 and 1, where:

- A value close to 1 implies a strong positive correlation.
- A value close to -1 implies a strong negative correlation.
- A value close to 0 implies little or no correlation.

Here are Pearson's correlation coefficients between HHI metrics:

• Traffic vs. Revenue: 0.055

This indicates a very weak positive correlation between traffic and revenue HHI. In other words, there is a slight tendency that when one increases, the other does too, but this relationship is very tenuous.

• Traffic vs. Accesses: -0.123

This indicates a weak negative correlation between traffic and access to HHI. This means that, generally, when the traffic HHI increases, the access HHI tends to decrease slightly, and vice versa.

• Revenue vs. Accesses: -0.773

This value indicates a moderate to strong negative correlation between the HHI of revenue and access. In other words, when the revenue HHI increases, the access HHI tends to decrease, and vice versa.

4.3. Correlation analysis

In analyzing the Colombian telecommunications market, it was noted that the relationship between traffic concentration and revenue is minimal, with a correlation of r = 0.055. This weak correlation suggests that the dynamics driving concentration in telecommunications resource usage (traffic) are not closely tied to those determining revenue concentration. For instance, operators with lesser traffic may have higher pricing strategies or premium services that allow them to maintain high revenues. On the other hand, the negative correlation between revenue and access, with a value of r = -0.773, is particularly telling. It indicates that companies dominating in revenue do not necessarily have the most access and vice versa. Such disparity could arise from various factors, such as differences in market strategies,

service diversification, and pricing structures. Recognizing these distinctions is crucial when formulating policies or making strategic decisions.

4.4. Volatility analysis

Volatility can be measured using the standard deviation, which measures the dispersion or variability of the data. A high standard deviation indicates that the data is scattered and therefore there is greater volatility. Conversely, a low standard deviation suggests that the data is more concentrated around the mean, indicating lower volatility. I calculate the standard deviation for the HHI of traffic, revenue, and accesses over time.

• Traffic Volatility: 230.40

This measure indicates that the traffic HHI has fluctuated, on average, by 230.40 points around its average during the period analyzed.

• Revenue Volatility: 314.02

The HHI of revenue has had, on average, a variation of 314.02 points around its average during the period analyzed.

• Access Volatility: 395.21

The HHI of accesses has fluctuated, on average, by 395.21 points around its average during the period analyzed.

These measures suggest that access to HHI has been the most volatile over time, followed by revenue and traffic. A more volatile HHI indicates an ever-changing market, while a more stable HHI suggests a more stationary market. The increased volatility in access to HHI could reflect a dynamic market environment regarding network access.

The volatility analysis uncovers substantial differences across the assessed metrics. The standard deviation in traffic stands at 230.40, in revenue it's 314.02, and in access, it peaks with a value of 395.21. These figures point towards a particularly fluid market dynamic concerning access. Such volatility may indicate constant competitive shifts, with operators rapidly adapting to market conditions or rolling out innovative strategies in terms of network accessibility. In contrast, while traffic and revenue have fluctuated, their volatility has decreased. This might highlight a consistency in dominant positions within these areas, where, despite market upheavals, certain firms sustain a steadfast presence, yielding steady revenue and traffic over time.

5. Conclusions

The measurement and monitoring of competition in the telecommunications market allow the state to implement strategies to improve market regulation to avoid the formation of oligopolies and monopolies.

The HHI index is a good index for measuring market concentration in the sense that it considers the relative size of each company in the market. Two companies with the same market share are not the same as one with 70% and another with 30%.

The LSTM model performs well in analyzing and projecting a time series' behavior. As an added value, it can identify more complex patterns in the data than other models, such as time series.

The gradual decline in the HHI in traffic and access suggests slightly increasing competition. However, revenue concentration has been more stable, slightly increasing in the middle of the decade. These trends could be related to operators' entry or exit from the market, regulatory changes, or technological innovations that alter market dynamics.

Future work aims to develop a hybrid model combining the results of at least three of the most widely used concentration indices to better predict concentration behavior in the telecommunications market. This will allow for the early proposal of effective strategies and methodologies to improve competition in Colombian telecommunications service operators.

Author contributions: Conceptualization, CH; methodology, CH, FM and LP; software, FM; validation, CH, FM and LP; formal analysis, FM and LP; data curation, FM and LP; writing—original draft preparation, CH, FM and LP; writing—review and editing, CH, FM and LP; project administration, CH; funding acquisition, YY. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministerio de Ciencia, Innovación y Tecnología (MinCiencias), Comisión de Regulación de las Comunicaciones and Universidad Distrital Francisco José de Caldas; grant number 082-2022.

Acknowledgments: The authors thank the Investigations Office (ODI) of the Universidad Distrital Francisco Jose de Caldas and Minciencias for supporting this investigative project.

Conflict of interest: The authors declare no conflict of interest.

References

- Amin, A., Shah, B., Khattak, A. B., et al. (2018). Just-in-time customer churn prediction: With and without data transformation. In: Proceeding of the 2018 IEEE congress on evolutionary computation (CEC). pp. 1-6.
- Apolinario, Y. B., Quispe, D. K. P., Rodriguez, P. B. V., et al. (2023). Bank concentration, measured by different indicators. the peruvian case. J. Globalization, Competitiveness, Governability, 16(1), 98–114. https://doi.org/10.3232/gcg.2022.v16.n1.05
- Bardey, D., Aristizábal, D., Gómez, J. S., et al. (2022). Concentration of the mobile telecommunications markets and countries' competitiveness. Telecommunications Policy, 46(1), 102230. https://doi.org/10.1016/j.telpol.2021.102230
- Berr Berradi, Z., Lazaar, M., Mahboub, O., et al. (2020). A Comprehensive Review of Artificial Intelligence Techniques in Financial Market. 2020 6th IEEE Congress on Information Science and Technology (CiSt), 132, 367–371. https://doi.org/10.1109/cist49399.2021.9357175
- Canadian Radio-television and Telecommunications Commission. (2023). Current trends—Mobile wireless. Available online: https://crtc.gc.ca/eng/publications/reports/PolicyMonitoring/mob.htm (accessed on 27 May 2023).
- Commission for the Regulation of Communications. (2022a). Battery of indicators for competition analysis | Postscript. Available online: https://postdata.gov.co/story/bateria-de-indicadores-para-el-analisis-de-competencia. doi 10.1111/ecaf.12498 (accessed on 12 July 2023).
- Commission for the Regulation of Communications. (2022b). CRC Regulatory Agenda 2021–2022. Available online: https://www.crcom.gov.co/sites/default/files/agenda/201229_ar_2021-22_vpub_0.pdf. doi 10.3233/IP-211532 (accessed on 27 May 2023).
- Commission for the Regulation of Communications. (2023a). Beyond the data. Available online: https://postdata.gov.co/. doi 10.3402/ejpt.v4i0.21502 (accessed 24 July 2023).
- Commission for the Regulation of Communications. (2023b). Industry Report of the ICT and Postal Sectors 2019. Available online: https://www.postdata.gov.co/sites/default/files/Reporteticpostal2019.pdf (accessed on 27 May 2023).

- Fiedler, M., De Moor, K., Ravuri, H., et al. (2017). Users on the move: On relationships between QoE ratings, data volumes and intentions to churn. In: Proceeding of the 42nd Conference on Local Computer Networks Workshops (LCN Workshops).
- Ilchenko, M., Uryvsky, L., Osypchuk, S. (2019). World trends of modern information and telecommunication technologies development. In: Proceeding of the International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo).
- Lin, X, Lv, T. J., Chen, X. (2018). The coevolutionary relationship of technology, market and government regulation in telecommunications. China Communications, 15(8), 152-173.
- Lis-Gutiérrez, J. P. (2023). Concentration measures and market stability: An application for excel (market concentration and market stability measures: An application for excel. SSRN Electronic. https://doi.org/10.2139/ssrn.2279769
- Mehrotra, A., Menon, S. (2021). Telecommunication & Networking Changing Customer Profile & Preferences. In: Proceeding of the 2nd International Conference on Computation. Automation and Knowledge Management. (ICCAKM).
- Melnik, A., Shy, O., Stenbacka, R. (2008). Assessing market dominance. Journal of Economic Behavior and Organization, 68(1), 63–72. https://doi.org/10.1016/j.jebo.2008.03.010
- Miera Berglind, O. S. (2015). Spectrum concentration and market competition. Implications for the use of caps in Mexico. In: Proceeding of the 2015 Conference of Telecommunication, Media and Internet Techno-Economics (CTTE).
- MinTIC. (2024). Quarterly ICT Bulletin. Available online: https://colombiatic.mintic.gov.co/679/articles-338221_archivo_pdf.pdf. doi 10.3926/jotse.314 (accessed 10 May 2024).
- Nekmahmud, M. D., Rahman, S. (2018). Measuring the Competitiveness Factors in Telecommunication Markets. In: Competitiveness in Emerging Markets: Market Dynamics in the Age of Disruptive Technologies. Cham: Springer International Publishing. pp. 339-372. https://10.1007/978-3-319-71722-718.
- OCDE.
 (2023).
 Informe:
 Economy
 Profile
 of
 Colombia
 2020.
 Available
 online:

 https://www.doingbusiness.org/content/dam/doingBusiness/country/c/colombia/COL.pdf.
 doi
 10.1080/01495933.2011.545686 (accessed on 27 May 2023).
 doi

OFCOM. (2022). Telecommunications Market Data Update Q3. OFCOM.

Ruíz Porras, A., Hernández Reyes, J. P. (2023). Concentration in banking markets and the technical efficiency of commercial banks in Mexico. Acta Universitaria, 30, 1–23. https://doi.org/10.15174/au.2020.2757