

Análisis Multivariado y Aprendizaje Automático en la evaluación y pronóstico de los perfiles financieros en el sector de Telecomunicaciones en Colombia

Multivariate Analysis and Machine Learning in the evaluation and forecasting of financial profiles in the telecommunications sector in Colombia

Efrain De La Hoz Granadillo ¹, Tomás Fontalvo Herrera ², Enrique De La Hoz Dominguez ³

¹ Ph.D en Ingeniería Industrial, Universidad de Cartagena, Grupo de investigación GOINCO, Cartagena, Colombia

² Ph.D en Administración, Universidad de Cartagena, Grupo de investigación CYPROI, Cartagena, Colombia

³ M.Sc en Ingeniería Industrial, Universidad Tecnológica de Bolívar, Cartagena, Colombia

Cite this article as: E. De La Hoz, T. Fontalvo, E. De La Hoz-Domínguez "Análisis Multivariado y Aprendizaje Automático en la evaluación y pronóstico de los perfiles financieros en el sector de Telecomunicaciones en Colombia", *Prospectiva*, Vol 18, N° 1, 7-12, 2020.

Recibido: 01/07/2019 / Aceptado: 01/11/2019

<http://doi.org/10.15665/rp.v18i1.2063>

RESUMEN

En esta investigación se desarrolla un método apoyado en las técnicas de Aprendizaje automático para evaluar y pronosticar perfiles financieros en el sector de Telecomunicaciones en Colombia. Lo anterior soportado en elementos conceptuales relacionados con el Análisis multivariado y herramientas de Aprendizaje automático. Para lo anterior, se utilizó información relacionada con los rubros financieros de 75 empresas, información que sirvió para calcular indicadores financieros y de productividad. Seguidamente se aplicó la técnica de análisis de conglomerados que permitió identificar y clasificar las empresas en tres grupos característicos del sector lográndose un nivel de homogeneidad de 0,527 y heterogeneidad de 1,358. A partir de lo anterior se aplicó el algoritmo GLMNET asociado a las técnicas de Aprendizaje automático, lográndose un modelo que predice de manera correcta la pertenencia a los grupos identificados con un 98% de precisión. En general se valora el método que integra el análisis de conglomerados y el algoritmo de GLMNET para evaluar y pronosticar perfiles financieros y de productividad en el sector de telecomunicaciones en Colombia.

Palabras Clave: Aprendizaje automático, Análisis multivariado, Análisis de Conglomerados, Indicadores financieros

ABSTRACT

In this research, a method based on the techniques of Machine learning is developed to evaluate and forecast financial profiles in the Telecommunications sector in Colombia. The foregoing is supported by conceptual elements related to multivariate analysis and Machine learning tools. For the above, information related to the financial rubrics of 75 companies was used, information that was used to calculate financial and productivity indicators. The conglomerate analysis technique was applied to identify and classify the companies in three characteristic groups of the sector, achieving a level of homogeneity of 0,527 and heterogeneity of 1.358. Based on the above, the GLMNET algorithm associated with the techniques of Machine learning was applied, achieving a model that correctly predicts belonging to the identified groups with 98% accuracy. In general, the method that integrates the cluster analysis and the GLMNET algorithm is evaluated to evaluate and forecast financial and productivity profiles in the telecommunications sector in Colombia.

Key words: Machine learning, Multivariate analysis, Conglomerate analysis, Financial indicators

1. INTRODUCCIÓN

El rápido desarrollo de la tecnología obliga a las empresas del sector de telecomunicaciones a estar en constante innovación de sus productos y servicios de manera que se pueda responder al dinamismo del sector a nivel mundial [1]. Particularmente en Colombia, el sector de telecomunicaciones es cada vez más competido [2], con costos fijos altos y caracterizado por la alta imitación de ofertas. Igualmente se viene generando un mayor uso de los servicios de telecomunicaciones con muy buenos resultados para las empresas.

De esta manera, la combinación de riesgos a corto plazo y presiones a largo plazo, a las cuales se encuentra expuesta el sector, conmina a que estas actúen de manera proactiva, actualizando y ampliando su propuesta de valor a la vez que se gestionan los riesgos a los que se enfrenta el sector. De esta manera, los riesgos más relevantes a los que enfrentan las empresas de telecomunicaciones se pueden agrupar en: acceso a la financiación, aumento de la competencia, irrupción de nuevas tecnologías, variaciones en el tipo de cambio y presión de los tipos de interés [3]. Así mismo, el desarrollo de la tecnología, ha contribuido en gran manera en el crecimiento y desarrollo del sector de las telecomunicaciones lo cual ha impactado positivamente en el mejoramiento de las empresas dedicadas al suministro de bienes y servicio al facilitar la comunicación y el flujo de información [4]. Así mismo, el proceso de globalización de las economías ha puesto de manifiesto la importancia de las telecomunicaciones observándose aplicaciones que han cambiado la forma en que estas hacen sus negocios [5]

En el presente trabajo de investigación, se busca desarrollar una metodología que identifique perfiles característicos del sector de telecomunicaciones de Colombia a partir de los indicadores financieros Margen Bruto, Margen operacional y Margen neto y de los indicadores de productividad Razón utilidad bruta y valor agregado, Razón utilidad operacional y valor agregado, Razón utilidad neta y valor agregado, Productividad del capital, Razón utilidad operativa y capital y Razón utilidad neta y capital. Para lo cual, se analizaron 75 empresas del sector. Por lo anterior en esta investigación se le da respuesta a las siguientes preguntas problemas ¿Cómo establecer los criterios para realizar un análisis y proyección de perfiles característicos de las empresas de telecomunicaciones de Colombia?, ¿Cuáles son los indicadores financieros y de productividad para desarrollar grupos característicos en el sector de telecomunicaciones?, ¿Cómo clasificar los resultados financieros y de competitividad en empresas del sector de telecomunicaciones de Colombia?, ¿Qué herramientas de Machine Learning permiten proyectar la pertenencia a perfiles de resultados financieros y de productividad empresarial?

2. MARCO TEÓRICO

Diversos trabajos de investigación muestran las ventajas y bondades de las técnicas de aprendizaje automático o Machine Learning para identificar patrones y predecir los resultados y comportamiento de un sistema proceso [6]–[9]. En este sentido, la técnica de aprendizaje automático GLMNET incorpora un algoritmo del tipo ensamble el cual se desarrolla a partir de un modelo lineal generalizado, el modelo de regresión de Ridge [10] y la regresión de Lasso [11]. Para el proceso de enlace se penaliza la magnitud y el número de coeficientes del modelo final de regresión, con lo cual se supera el riesgo de sobre ajuste del modelo. Estas técnicas, han mostrado ser altamente efectivas en conjuntos de datos con muchas variables predictoras y pocas observaciones individuales. Así, en los casos límites en el que el parámetro de enlace asume un valor de cero, el modelo se desarrolla como una regresión de Ridge y en los casos que toma un valor de 1 se desarrolla como una regresión de Lasso, en otros casos como, por ejemplo, un valor de $\Lambda = 0,05$ tendrá como resultado un modelo con un 95% de regresión de Ridge y un 5% de regresión de Lasso. Esta penalización es muy importante en los casos en los que el conjunto de datos es pequeño o en los casos de que existan muchas variables predictoras correlacionadas.

El algoritmo GLMNET [12], define una variable de respuesta a partir de un vector de variables predictoras $X \in R^p$, aproximando las funciones de regresión por el modelo lineal $E(Y | X = x) = \beta_0 + X^T \beta$, así se cuenta con N parejas de observaciones (X_i, Y_i) , resolviendo el problema de optimización planteado en la ecuación (1)

$$\min_{(\beta_0, \beta) \in R^{p+1}} \quad (1)$$

Lo anterior es el sustento para el proceso de pronóstico de las empresas en el objeto de estudio de esta investigación Como criterios de validez se utilizará la matriz de confusión, el análisis de sensibilidad y especificidad

Así mismo, el Análisis Multivariado ha mostrado su importancia en el análisis de problemas en los que intervienen un gran número de variables [13]–[15], representa un conjunto de herramientas estadísticas y matemáticas utilizadas para analizar e interpretar los datos asociadas a múltiples variables [16]. Tiene como finalidad analizar de manera simultánea un conjunto de datos multivariantes medidos para cada individuo u objeto estudiado, lo que permite una mejor comprensión del fenómeno o proceso objeto de estudio [16]. Dentro de las técnicas multivariadas, el análisis de conglomerados se caracteriza por agrupar elementos o variables, buscando lograr la mayor homogeneidad posible en cada grupo y la máxima diferencia entre los mismos. De la Fuente en [4], su importancia radica en su aplicabilidad en muchas áreas de

investigación, para desarrollar análisis descriptivo. Así mismo la caracterización de perfiles financieros y de productividad mediante la aplicación de técnicas multivariantes, en el sector de telecomunicaciones en Colombia, es importante para su evaluación de severa importancia, al momento de llevar estudiar el comportamiento financiero de las empresas de telecomunicaciones del país. [19]. La aplicación del análisis multivariado, permite identificar grupos característicos del sector analizado, con lo cual se identifican oportunidades de mejora y orienta en el diseño de políticas públicas para el desarrollo del sector de telecomunicaciones.

Para el desarrollo de la investigación se utilizaron indicadores financieros y de productividad, los cuales permiten el análisis de la realidad financiera y de productividad de manera individual, y facilitan la comparación de la misma con la competencia y con la entidad u organización que lidera. Por lo que se llevará a cabo un estudio y análisis de las empresas prestadoras de este servicio, tomando de cada una de ellas indicadores financieros Margen Bruto, Margen operacional y Margen neto, así como los indicadores de productividad Razón utilidad bruta y valor agregado, Razón utilidad operacional y valor agregado, Razón utilidad neta y valor agregado, Productividad del capital, Razón utilidad operativa y capital y Razón utilidad neta y capital, los cuales permitieron analizar el sector de telecomunicaciones a partir de sus resultados.

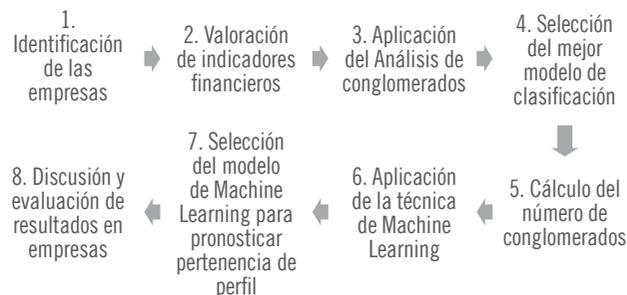
3. METODOLOGÍA

Para el desarrollo de la investigación, se tomó información de 75 empresas del sector de telecomunicaciones que presentaron sus estados financieros en el Sistema de Información y Reporte Empresarial - SIREM de la Superintendencia de Sociedades. Con los datos asociados al Balance General y el estado de resultados, se determinaron los indicadores financieros Margen Bruto, Margen operacional y Margen neto, así como los indicadores de productividad Razón utilidad bruta y valor agregado, Razón utilidad operacional y valor agregado, Razón utilidad neta y valor agregado, Productividad del capital, Razón utilidad operativa y capital y Razón utilidad neta y capital. Con los indicadores anteriores, se desarrolló un análisis de clúster que permitió identificar grupos o perfiles característicos de las empresas. Seguidamente, se aplicaron técnicas de Machine Learning para desarrollar un modelo multivariado con el cual pronosticar y predecir la pertenencia de una empresa a los grupos o perfiles empresariales característicos del sector. En la figura 1, se presenta de manera esquemática el método seguido en el desarrollo de la investigación.

Para el análisis de clúster, se utilizó el software IBM SPSS Statistics 25, lo cual implicó un análisis de homogeneidad utilizando medidas de distancia promedio y máxima desde el

centroide, un análisis de heterogeneidad utilizando la medida entre centroide y el gráfico de dendograma que junto al análisis de la gráfica de codo realizado con las distancias de aglomeración permitieron definir el método de medida de distancia, método de aglomeración y número de grupos apropiados para caracterizar el sector. En la aplicación de la técnica de Machine Learning, se utilizó el software Weka explorer, soportado en el algoritmo de GLMNET, la matriz de confusión y el análisis de sensibilidad y especificidad, con lo cual se estableció la estructura para pronosticar la pertenencia de empresas en los 3 niveles de resultados financieros y productividad.

Figura 1. Metodología para valorar, clasificar y predecir empresas eficientes y no eficientes
Figure 1. Methodology to assess, classify and predict efficient and inefficient companies



4. RESULTADOS Y DISCUSIÓN

Con la información asociada a los estados financieros de 75 empresas del sector de telecomunicaciones de Colombia, se calcularon los indicadores financieros (Margen bruto -MB, Margen operacional - Margen neto- MN) y de productividad (Razón utilidad bruta y valor agregado- RUB-VA, Razón utilidad operacional y valor agregado RUO-VA, Razón utilidad neta y valor agregado -RUN-VA, Productividad del capital -PC, Razón utilidad operativa y capital de trabajo RUO-CT y Razón utilidad neta y capital de trabajo -RUN-CT). Con los indicadores anteriores, se desarrolló un análisis de conglomerados, los resultados muestran el mejor comportamiento de agrupación utilizando el criterio enlace de Ward y la medida de distancia Euclidiana en la figura 2, se presenta el gráfico de dendograma en el que se evidencia la conveniencia de usar de grupos de conglomerados.

Así mismo en la Tabla 1, se presentan los resultados del análisis de homogeneidad, en el que se observan los tres conglomerados con 36, 19 y 20 observaciones en cada uno, destacándose bajos valores de distancia promedio en cada conglomerado (0,543; 0,676; 0,364) los cuales son inferiores a las distancias registradas entre conglomerados en el análisis de heterogeneidad de la Tabla 2, verificándose analíticamente la decisión de conformar los tres niveles o grupos de empresas asociado a sus resultados en los indicadores financieros y de productividad.

Figura 2. Dendrograma
Figure 2. Dendrogram



Tabla 1. Análisis de Homogeneidad de Conglomerados
Table 1. Conglomerate Homogeneity Analysis

	Número de observaciones	Dentro de la suma de cuadrados del conglomerado	Distancia promedio desde el centroide	Distancia máxima desde centroide
Conglomerado1	36	13,3136	0,543	1,240
Conglomerado2	19	10,3502	0,676	1,315
Conglomerado3	20	2,8714	0,364	0,507

Tabla 2. Análisis de Heterogeneidad de Conglomerados
Table 2. Analysis of Heterogeneity of Clusters

	Conglomerado1	Conglomerado2	Conglomerado3
Conglomerado1	0,00000	2,01837	0,95107
Conglomerado2	2,01837	0,00000	1,10512
Conglomerado3	0,95107	1,10512	0,00000

En la Tabla 3, se presenta el análisis de media por conglomerado que caracteriza cada uno de los tres conglomerados

Tabla 3. Análisis de media de Conglomerados
Table 3. Conglomerate average analysis

Variable	Conglomerado1	Conglomerado2	Conglomerado3	Centroide principal
MB (%)	0,42841	0,35023	0,47811	0,42186
MO (%)	0,06223	0,04844	0,10348	0,06973
MN (%)	0,00662	0,00196	0,04923	0,01680
RUB-VA	0,40699	0,34229	0,46239	0,40537
RUO-VA	0,05549	0,04829	0,10054	0,06568
RUN-VA	0,00125	0,00392	0,04853	0,01454
PC (%)	1,04194	3,05610	1,97378	1,80068
RUO-CT	0,07179	0,15190	0,19920	0,12606
RUN-CT	0,01401	0,01265	0,09558	0,03542

Seguidamente, teniendo en cuenta la clasificación de los tres clústeres o grupos de empresas asociado a sus resultados en los indicadores financieros y de productividad del sector de telecomunicaciones, se aplicó el modelo de Minería de datos Machine Learning GLMNET, ampliamente utilizado para el análisis de poblaciones pequeñas, con el cual se desarrolló el proceso de pronóstico de pertenencia a los conglomerados 1, 2 y 3 respectivamente. En el análisis se utilizó la técnica de validación cruzada con grupos aleatorios de tamaño 10, se tomó como variable de predicción el clúster o conglomerado y variables predictoras Margen Bruto, Margen operacional, Margen neto, Razón utilidad bruta y valor agregado, Razón utilidad operacional y valor agregado, Razón utilidad neta y valor agregado, Productividad del capital, Razón utilidad operativa y capital y Razón utilidad neta y capital. Lo anterior permitió estructurar la metodología de evaluación y pronóstico mediante técnicas de Machine Learning.

Los resultados del algoritmo GLMNET, generó una precisión en el pronóstico de la pertenencia a los grupos de conglomerado del 94,66% lo que evidencia una alta capacidad de predecir la pertenencia a partir de las variables predictoras analizadas. La Tabla 4, muestra la matriz de confusión del proceso de clasificación y pronóstico

Tabla 4. Matriz de confusión algoritmo GLMNET
Table 4. GLMNET algorithm confusion matrix

Predicción	Grupo	Precisión 94,66%			Total
		Conglomerado 1	Conglomerado 2	Conglomerado 3	
Conglomerado 1		34	1	1	36
Conglomerado 2		0	19	0	19
Conglomerado 3		2	0	18	20

Estos resultados son significativos teniendo en cuenta el tamaño reducido de la población analizada. Con el propósito de valorar y comparar el resultado anterior, se procedió a aplicar de manera complementaria el algoritmo Random Forest, los resultados de la Tabla 5 muestran la capacidad de predicción correcta del 93,33% inferior al del algoritmo GLMNET.

Tabla 5. Matriz de confusión algoritmo Random Forest
Table 5. Random Forest algorithm confusion matrix

Variable	Conglomerado1	Conglomerado2	Conglomerado3	Centroide principal
MB (%)	0,42841	0,35023	0,47811	0,42186
MO (%)	0,06223	0,04844	0,10348	0,06973
MN (%)	0,00662	0,00196	0,04923	0,01680
RUB-VA	0,40699	0,34229	0,46239	0,40537

Complementario al análisis de precisión del algoritmo de GLMNET para realizar pronóstico, se analizó el poder de discriminación por conglomerado, los resultados muestran un valor de sensibilidad del 93,33% (capacidad de clasificación de grupos) y de especificidad del 93,6% (Capacidad de clasificación específica) de donde se valora la capacidad del modelo para identificar grupos y unidades específicas. Los resultados de esta investigación superan trabajos como los de [17]–[20] en los que las precisiones del pronóstico de clasificación son menores.

5. CONCLUSIONES

De manera general en el presente trabajo de investigación se desarrolla una metodología que articula técnicas de clasificación y pronóstico, que aporta en los procesos de análisis de variables y resultados de la gestión aplicados en el contexto del sector de telecomunicaciones, la cual se sintetiza en la figura 1. De lo anterior se concluye la generación de nuevo conocimiento toda vez que se integra la técnica de análisis de conglomerado con el algoritmo GLMNET para identificar y predecir la pertenencia de empresas a grupos característicos homogéneos intra-grupo y heterogéneos extra-grupo. Y en función de esto mejorar la comprensión, análisis, toma de decisiones y la mejora asociada a sus resultados financieros y de productividad. Así mismo, como resultado de este trabajo, se evidencia la robustez del algoritmo de GLMNET para pronosticar la pertenencia a grupos característicos de las empresas pertenecientes a sectores empresariales, lográndose un nivel del 98% de precisión en la clasificación correcta, es decir a partir de la información asociada a los indicadores financieros Margen Bruto, Margen operacional y Margen neto, y de los indicadores de productividad Razón utilidad bruta

y valor agregado, Razón utilidad operacional y valor agregado, Razón utilidad neta y valor agregado, Productividad del capital, Razón utilidad operativa y capital y Razón utilidad neta y capital. Como líneas futuras de investigación, se invita a la comunidad científica a extender la aplicación de esta metodología en otros sectores empresariales evaluando otras variables de gestión.

REFERENCIAS

- [1] G. R. Zambrano and A. C. Gamboa, "Los procesos de investigación y desarrollo de las empresas operadoras de telecomunicaciones del Ecuador The processes of research and development of the Ecuador telecommunications operating companies," *Rev. Cuba. Ciencias Informáticas*, vol. 10, no. 4, pp. 15–27, 2016.
- [2] J. Algarra, *Los 5 grandes retos del sector de las telecomunicaciones*, Corporación. 2018.
- [3] J. Torrejón-Flores and F. García-Solanes, "Las variaciones del tipo de cambio y el índice de inflación en las economías emergentes," *Rev. CEPAL*, no. 116, pp. 27–46, 2015.
- [4] R. Orduz, L. Ayala, and M.-E. Vallejo, *Aprender y Educar con las tecnologías del siglo XXI*, Colombia D. Bogotá-Colombia, 2012.
- [5] J. Díaz Lazo, A. Pérez Gutiérrez, and R. Florido Bacallao, "Impacto de las tecnologías de la información y las comunicaciones (TIC) para disminuir la brecha digital en la sociedad actual impact of information technology and communications (ICT) to reduce the digital divide in today's society," *Cultiv. trópicales*, vol. 32, no. 1, pp. 81–90, 2011.
- [6] B. Journal, J. Fischer, V. S. Lopes, S. L. Cardoso, U. C. Filho, and V. L. Cardoso, "Machine learning techniques applied to lignocellulosic ethanol in simultaneous hydrolysis and fermentation," *Brazilian J. Chem. Eng.*, vol. 34, no. 01, pp. 53–63, 2017.
- [7] R. Piazzaroli, "An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements," *J. Solids Struct.*, vol. 16, no. 1998, pp. 1–17, 2019.
- [8] A. Cient, A. M. Guimar, K. Sato, and C. Ferreira, "Estimating gypsum equipment under no-till based on machine," *Rev. Ciência Agronômica*, vol. 6, no. 2, pp. 250–257, 2015.
- [9] F. Silva, V. Vidotti, F. Cremasco, M. Dias, E. Gomi, and V. Costa, "Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry," *Arq. Bras. Oftalmol.*, vol. 76, no. 3, pp. 170–174, 2013.
- [10] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [11] C. Hans, "Bayesian lasso regression," *Biometrika*, vol. 96, no. 4, pp. 835–845, 2009.
- [12] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Stat. software*, vol. 33, no. 1, pp. 1–22, 2010.
- [13] C. Pereira *et al.*, "Multivariate Analysis and Machine Learning in Properties of Ultisols (Argissolos) of Brazilian Amazon," *Rev Bras Cienc Solo*, vol. 42, no. e0170419, pp. 1–20, 2018.
- [14] S. Silva and J. De Souza, "Multivariate Analysis and Geostatistics of the Fertility of a Humic Rhodic Hapludox under," *Rev Bras Cienc Solo. Bras. Ci. Solo*, vol. 36, no. 2, pp. 467–474, 2012.
- [15] C. Moreno, I. Mancebo, A. M. Tarquis, and M. M. Moreno, "Uni-

variate and multivariate analysis on processing tomato quality under different mulches,” *Sci. Agric.*, vol. 71, no. 2, pp. 114–119, 2014.

- [16] C. Cuadras, *Nuevos métodos de análisis multivariante*, CMC Editio. Barcelona, 2019.
- [17] T. Fontalvo, E. De La Hoz, and E. De La Hoz, “Método Análisis Envolvente de Datos y Redes Neuronales en la Evaluación y Predicción de la Eficiencia Técnica de Pequeñas Empresas Exportadoras Data Envelopment Analysis Method and Neural Networks in the Evaluation and Prediction of the Technical Efficiency,” *Inf. Tecnol.*, vol. 29, no. 6, pp. 267–276, 2018.
- [18] J. Morelos, E. De La Hoz, and T. Fontalvo, “Método de cálculo multivariante para analizar y proyectar el comportamiento de las razones financieras de grupos empresariales del sector extracción en Colombia,” *Interciencia*, vol. 43, no. 10, pp. 696–700, 2018.
- [19] E. J. de la Hoz Granadillo, T. J. F. Herrera, and J. M. Gómez, “Evaluación del comportamiento de los indicadores de productividad y rentabilidad financiera del sector petróleo y gas en Colombia mediante el análisis discriminante,” *Contaduría y Adm.*, vol. 59, no. 4, pp. 167–191, 2014.
- [20] E. De La Hoz and L. López Polo, “Aplicación de Técnicas de Análisis de Conglomerados y Redes Neuronales Artificiales en la Evaluación del Potencial Exportador de una Empresa,” *Inf. tecnológica*, vol. 28, no. 4, pp. 67–74, 2017.